

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:21:09 ; Search time 53.6471 Seconds
(without alignments)
400.276 Million cell updates/sec

Title: US-09-620-955B-11
Perfect score: 372
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 76

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		8				
Result		Query				
No.	Score	Match	Length	DB	ID	Description
1	303	81.5	145	4	AAB69614	Aab69614 Huntingti
2	300.5	80.8	98	4	AAB69610	Aab69610 Huntingti
3	293	78.8	64	4	AAB69607	Aab69607 Huntingti
4	293	78.8	86	2	AAW95073	Aaw95073 GST-HD fu
5	293	78.8	86	2	AAW95078	Aaw95078 GST-HD fu
6	293	78.8	89	4	AAB69608	Aab69608 Huntingti
7	293	78.8	94	2	AAW95075	Aaw95075 GST-HD fu
8	293	78.8	94	2	AAW95080	Aaw95080 GST-HD fu
9	293	78.8	121	4	AAB69609	Aab69609 Huntingti

10	293	78.8	123	4	AAB69611	Aab69611	Huntingti
11	293	78.8	155	4	AAB69612	Aab69612	Huntingti
12	293	78.8	171	5	AAE26650	Aae26650	Human hun
13	268.5	72.2	59	4	AAB69605	Aab69605	Huntingti
14	255	68.5	108	2	AAW95071	Aaw95071	Amino aci
15	255	68.5	108	2	AAW95076	Aaw95076	Amino aci
16	219	58.9	722	7	ADC55550	Adc55550	Mutant hu
17	219	58.9	951	7	ADC55544	Adc55544	Mutant hu
18	218	58.6	66	4	AAB69613	Aab69613	Huntingti
19	217	58.3	149	4	AAB72673	Aab72673	Polygluta
20	216	58.1	113	4	AAB69615	Aab69615	Huntingti
21	214	57.5	128	5	ABP53102	Abp53102	Human Zac
22	214	57.5	134	8	ADE52654	Ade52654	Murine dn
23	213	57.3	905	5	ABG93053	Abg93053	S. cerevi
24	213	57.3	905	6	ABR53130	Abr53130	Protein s
25	211.5	56.9	338	5	AAU77921	Aau77921	Human Tat
26	211.5	56.9	371	2	AAW73369	Aaw73369	Epitope t
27	210	56.5	611	5	ABG93274	Abg93274	C. albica
28	207.5	55.8	325	4	ABB71488	Abb71488	Drosophil
29	207	55.6	910	2	AAZ22191	Aay22191	Mouse bra
30	207	55.6	910	5	ABJ10802	Abj10802	Mouse HCN
31	206.5	55.5	1024	6	AAG79909	Aag79909	MECT1-MAM
32	206.5	55.5	1153	6	AAG79910	Aag79910	MAML2. 4/
33	206.5	55.5	1153	7	ADC59310	Adc59310	Human pol
34	206	55.4	80	4	AAB69622	Aab69622	TATA bind
35	206	55.4	821	4	ABB70509	Abb70509	Drosophil
36	205	55.1	63	5	AAE26651	Aae26651	Human hun
37	205	55.1	3223	4	ABB11407	Abb11407	Human Hun
38	205	55.1	3223	4	ABB11470	Abb11470	Human Hun
39	204.5	55.0	1761	4	ABB59512	Abb59512	Drosophil
40	204	54.8	79	4	AAB69616	Aab69616	Huntingto
41	204	54.8	171	2	AAW99022	Aaw99022	Human hun
42	204	54.8	513	2	AAZ33500	Aay33500	Human hun
43	204	54.8	530	2	AAZ33501	Aay33501	Human apo
44	204	54.8	552	2	AAZ33502	Aay33502	Human apo
45	204	54.8	589	2	AAZ33503	Aay33503	Human apo

ALIGNMENTS

RESULT 1

AAB69614

ID AAB69614 standard; protein; 145 AA.

XX

AC AAB69614;

XX

DT 30-APR-2001 (first entry)

XX

DE Huntingtin accumulation inhibitor peptide GFP-HD-Q104.

XX

KW Neurological disorder; Huntington's disease; Alzheimer's disease;

KW Parkinson's disease; prion disease; frontotemporal dementia;

KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;

KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;

KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.

XX

RESULT 3

AAB69607

ID AAB69607 standard; protein; 64 AA.

XX

AC AAB69607;

XX

DT 30-APR-2001 (first entry)

XX

DE Huntingtin accumulation inhibitor peptide HD-Q47-GFP.

XX

KW Neurological disorder; Huntington's disease; Alzheimer's disease;

KW Parkinson's disease; prion disease; frontotemporal dementia;

KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;

KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;

KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.

XX

OS Synthetic.

XX

PN WO200106989-A2.

XX

PD 01-FEB-2001.

XX

PF 24-JUL-2000; 2000WO-US020131.

XX

PR 27-JUL-1999; 99US-0146047P.

PR 21-JUL-2000; 2000US-00620955.

XX

PA (HUST/) HUSTON J S.

PA (MESS/) MESSER A.

PA (LECE/) LECERF J.

XX

PI Huston JS, Messer A, Lecerf J;

XX

DR WPI; 2001-182700/18.

XX

PT Inhibiting intracellular polypeptide accumulation, useful for treating
PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT the polypeptide with a specific intrabody.

XX

PS Disclosure; Page 97; 108pp; English.

XX

CC The present invention describes a method for inhibiting the formation of
CC aggregates of certain proteins, involving contacting the protein with a
CC binding molecule known as an intrabody. Proteins to be bound include
CC those associated with neurological disorders, and so the method can be
CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC Huntington's diseases, prion disease, frontotemporal dementia,
CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7

XX

SQ Sequence 64 AA;

Query Match 78.8%; Score 293; DB 4; Length 64;
Best Local Similarity 98.4%; Pred. No. 7.6e-25;

Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	67	Q	67
Db	61	Q	61

RESULT 4

AAW95073

ID AAW95073 standard; protein; 86 AA.

XX

AC AAW95073;

XX

DT 20-MAY-1999 (first entry)

XX

DE GST-HD fusion protein GST-HD51DELP.

XX

KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;

KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;

KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;

KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;

KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.

XX

OS Synthetic.

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT Misc-difference 1

```
FT          /note= "this residue is connected to a GST protein which
FT          is not indicated in the sequence"
```

XX

PN WO9906838-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-EP004810.

XX

PR 01-AUG-1997; 97EP-00113320.

XX

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX

PI Wanker E, Lehrach H, Scherzinger E, Bates G:

XX

DR WPI; 1999-153955/13.

XX

PT Detecting amyloid-like fibrils or protein aggregates insoluble in

PT detergent or urea - from their retention on a filter, used for diagnosis,

PT particularly of diseases associated with polyglutamine expansion.

XX

PS Disclosure; Fig 8; 56pp; English.

XX

CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)

PN WO9906545-A2.
 XX
 PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98WO-EP004811.
 XX
 PR 01-AUG-1997; 97EP-00113306.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Wanker E, Lehrach H, Scherzinger E, Bates G;
 XX
 DR WPI; 1999-153775/13.
 XX
 PT Composition containing fusion protein that includes amyloidogenic peptide
 PT - able to self-assemble into fibrils or aggregates, used to detect and
 PT monitor neuronal diseases, and also to screen for therapeutic inhibitors.
 XX
 PS Disclosure; Fig 8; 62pp; English.
 XX
 CC The invention relates to a composition comprising a fusion protein of (i)
 CC (poly)peptide that increases solubility and/or prevents aggregation of
 CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
 CC assemble into amyloid-like fibrils or protein aggregates. Host cells
 CC transformed with a vector containing the nucleic acid encoding the fusion
 CC protein are used for the recombinant expression of the fusion protein.
 CC The composition is used to detect onset and progression of diseases
 CC associated with fibrils/protein aggregates. It is potentially useful for
 CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
 CC repeat expansion conditions such as Huntington's disease (HD), spinal and
 CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
 CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
 CC release of the amyloidogenic polypeptide from fusion protein have a
 CC precise starting time for aggregate formation, allowing kinetic
 CC measurements, and use of an enzyme for cleavage allows testing under
 CC physiological conditions. Sequences AAW95077-80 represent GST-HD fusion
 CC proteins
 XX
 SQ Sequence 86 AA;

Query Match 78.8%; Score 293; DB 2; Length 86;
 Best Local Similarity 98.4%; Pred. No. 1e-24;
 Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQQQL 66
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 8 MATLEKLMKAFESLKSFQQQ 67
 Qy 67 Q 67
 |
 Db 68 Q 68

RESULT 6
 AAB69608
 ID AAB69608 standard; protein; 89 AA.
 XX

AC AAB69608;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Huntingtin accumulation inhibitor peptide HD-Q72-GFP.
 XX
 KW Neurological disorder; Huntington's disease; Alzheimer's disease;
 KW Parkinson's disease; prion disease; frontotemporal dementia;
 KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
 KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
 KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
 XX
 OS Synthetic.
 XX
 PN WO200106989-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 24-JUL-2000; 2000WO-US020131.
 XX
 PR 27-JUL-1999; 99US-0146047P.
 PR 21-JUL-2000; 2000US-00620955.
 XX
 PA (HUST/) HUSTON J S.
 PA (MESS/) MESSER A.
 PA (LECE/) LECERF J.
 XX
 PI Huston JS, Messer A, Lecerf J;
 XX
 DR WPI; 2001-182700/18.
 XX
 PT Inhibiting intracellular polypeptide accumulation, useful for treating
 PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
 PT the polypeptide with a specific intrabody.
 XX
 PS Disclosure; Page 97; 108pp; English.
 XX
 CC The present invention describes a method for inhibiting the formation of
 CC aggregates of certain proteins, involving contacting the protein with a
 CC binding molecule known as an intrabody. Proteins to be bound include
 CC those associated with neurological disorders, and so the method can be
 CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
 CC Huntington's diseases, prion disease, frontotemporal dementia,
 CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
 CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
 CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
 XX
 SQ Sequence 89 AA;

Query Match 78.8%; Score 293; DB 4; Length 89;
 Best Local Similarity 98.4%; Pred. No. 1.1e-24;
 Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQQQL 66
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MATLEKLMKAFESLKSFQQ 60

Qy 67 Q 67
|
Db 61 Q 61

RESULT 7

AAW95075

ID AAW95075 standard; protein; 94 AA.

XX

AC AAW95075;

XX

DT 20-MAY-1999 (first entry)

XX

DE GST-HD fusion protein GST-HD51DELPBio.

XX

KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.

XX

OS Synthetic.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "this residue is connected to a GST protein which
FT is not indicated in the sequence"

XX

PN WO9906838-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-EP004810.

XX

PR 01-AUG-1997; 97EP-00113320.

XX

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX

PI Wanker E, Lehrach H, Scherzinger E, Bates G;

XX

DR WPI; 1999-153955/13.

XX

PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for diagnosis,
PT particularly of diseases associated with polyglutamine expansion.

XX

PS Disclosure; Fig 8; 56pp; English.

XX

CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for

CC treating the diseases. Other applications include detection of inclusion
 CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
 CC associated with polyglutamine expansion are particularly diagnosed, e.g.
 CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
 CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
 CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
 CC scrapie. The protein aggregates can now be detected simply, routinely and
 CC rapidly, without requiring sophisticated equipment. The method can be
 CC made quantitative, by analysing a series of dilutions, and can be
 CC automated to allow many samples to be analysed on the same filter.
 CC Sequences AAW95072-75 represent GST-HD fusion proteins

XX

SQ Sequence 94 AA;

Query Match 78.8%; Score 293; DB 2; Length 94;
 Best Local Similarity 98.4%; Pred. No. 1.1e-24;
 Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQAAA 66
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 8 MATLEKLMKAFESLKSFQAAA 67

Qy 67 Q 67
 |
 Db 68 Q 68

RESULT 8

AAW95080

ID AAW95080 standard; protein; 94 AA.

XX

AC AAW95080;

XX

DT 20-MAY-1999 (first entry)

XX

DE GST-HD fusion protein GST-HD51DELPBio.

XX

KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
 KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
 KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
 KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
 KW GST-HD; HD.

XX

OS Synthetic.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "this residue is connected to a GST protein which
 FT is not indicated in the sequence"

XX

PN WO9906545-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-EP004811.

XX

XX

XX

XX

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

cc
aa

CC

cc

CC

CC

CC

CC

CC

CC

CC

XX

SO

Query Match 78.8%; Score 293; DB 2; Length 94;

Best Local Similarity 98.4%; Pred. No. 1.1e-24;

Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 MATLEKLMKAFESLKSFQQQL 66

Db 8 MATLEKLMKAFESLKSFOO 67

Qy 67 Q 67

Db 68 0 68

RESULT 9

AAB69609

ID AAB69609 standard; protein; 121 AA.

XX

AC AAB69609;

XX

DT 30-APR-2001 (first entry)

XX

DE Huntingtin accumulation inhibitor peptide HD-O104-GFP.

XX

KW Neurological disorder; Huntington's disease; Alzheimer's disease;
 KW Parkinson's disease; prion disease; frontotemporal dementia;
 KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
 KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
 KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
 XX
 OS Synthetic.
 XX
 PN WO200106989-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 24-JUL-2000; 2000WO-US020131.
 XX
 PR 27-JUL-1999; 99US-0146047P.
 PR 21-JUL-2000; 2000US-00620955.
 XX
 PA (HUST/) HUSTON J S.
 PA (MESS/) MESSER A.
 PA (LECE/) LECERF J.
 XX
 PI Huston JS, Messer A, Lecerf J;
 XX
 DR WPI; 2001-182700/18.
 XX
 PT Inhibiting intracellular polypeptide accumulation, useful for treating
 PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
 PT the polypeptide with a specific intrabody.
 XX
 PS Disclosure; Page 98; 108pp; English.
 XX
 CC The present invention describes a method for inhibiting the formation of
 CC aggregates of certain proteins, involving contacting the protein with a
 CC binding molecule known as an intrabody. Proteins to be bound include
 CC those associated with neurological disorders, and so the method can be
 CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
 CC Huntington's diseases, prion disease, frontotemporal dementia,
 CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
 CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
 CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
 XX
 SQ Sequence 121 AA;

Query Match 78.8%; Score 293; DB 4; Length 121;
 Best Local Similarity 98.4%; Pred. No. 1.4e-24;
 Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQQQL 66
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MATLEKLMKAFESLKSFQQ 60

 Qy 67 Q 67
 |
 Db 61 Q 61

RESULT 10

AAB69611
 ID AAB69611 standard; protein; 123 AA.
 XX
 AC AAB69611;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Huntingtin accumulation inhibitor peptide HD-Q72-Myc-HIS6.
 XX
 KW Neurological disorder; Huntington's disease; Alzheimer's disease;
 KW Parkinson's disease; prion disease; frontotemporal dementia;
 KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
 KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;
 KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
 XX
 OS Synthetic.
 XX
 PN WO200106989-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 24-JUL-2000; 2000WO-US020131.
 XX
 PR 27-JUL-1999; 99US-0146047P.
 PR 21-JUL-2000; 2000US-00620955.
 XX
 PA (HUST/) HUSTON J S.
 PA (MESS/) MESSER A.
 PA (LECE/) LECERF J.
 XX
 PI Huston JS, Messer A, Lecerf J;
 XX
 DR WPI; 2001-182700/18.
 XX
 PT Inhibiting intracellular polypeptide accumulation, useful for treating
 PT neurological disorders, e.g. Alzheimer's disease, comprises contacting
 PT the polypeptide with a specific intrabody.
 XX
 PS Disclosure; Page 98-99; 108pp; English.
 XX
 CC The present invention describes a method for inhibiting the formation of
 CC aggregates of certain proteins, involving contacting the protein with a
 CC binding molecule known as an intrabody. Proteins to be bound include
 CC those associated with neurological disorders, and so the method can be
 CC used in the prevention of diseases such as Alzheimer's, Parkinson's and
 CC Huntington's diseases, prion disease, frontotemporal dementia,
 CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
 CC dentatorubal-pallidoluysian atrophy, spinocerebellar ataxia type 1
 CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7
 XX
 SQ Sequence 123 AA;

Query Match 78.8%; Score 293; DB 4; Length 123;
 Best Local Similarity 98.4%; Pred. No. 1.5e-24;
 Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 MATLEKLMKAFESLKSFQQQL 66

CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7

Query Match 78.8%; Score 293; DB 4; Length 155;
Best Local Similarity 98.4%; Pred. No. 1.8e-24;
Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 12

XX
 PS Disclosure; Page 88; 93pp; English.
 XX
 CC The present invention relates to novel screening methods for identifying
 CC therapeutic agents for diseases associated with protein misfolding. The
 CC method involves contacting a yeast cell with a candidate compound, where
 CC the yeast cell expresses a polypeptide comprising a misfolded disease
 CC protein, contacting the yeast cell with a toxicity inducing agent and
 CC evaluating the yeast cell for viability, where the viability indicates
 CC the candidate compound is a candidate therapeutic agent. The method is
 CC useful to screen for therapeutic agents for diseases associated with
 CC protein misfolding such as Alzheimer's disease (AD), Parkinson's disease
 CC (PD), Familial amyloid polyneuropathy, tauopathies (e.g. Pick disease,
 CC lobar atrophy, frontotemporal dementia) or trinucleotide diseases (e.g.
 CC Huntington's disease, spinocerebellar ataxia (SCA), fragile-X syndrome,
 CC myotonic dystrophy, dentatorubral pallidoluysian atrophy (DRPLA) and
 CC prion diseases (e.g. Creutzfeldt-Jacob disease (CJD), fatal familia
 CC insomnia (FFI), Gerstmann-Straussler-Scheinker disease (GSS), mad cow
 CC disease, scrapie and kuru). The method is useful for treating a patient
 CC with Huntington's disease or Parkinson's disease. The present sequence is
 CC human huntington (htQ103) protein. This sequence is used to illustrate
 CC the method of the invention
 XX
 SQ Sequence 171 AA;

Query Match 78.8%; Score 293; DB 5; Length 171;
 Best Local Similarity 98.4%; Pred. No. 2e-24;
 Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQAAA 66
 |||
 Db 1 MATLEKLMKAFESLKSFQAAA 60
 Qy 67 Q 67
 |
 Db 61 Q 61

RESULT 13

AAB69605

ID AAB69605 standard; protein; 59 AA.

XX

AC AAB69605;

XX

DT 30-APR-2001 (first entry)

XX

DE Huntingtin accumulation inhibitor peptide GST-HD-Q25.

XX

KW Neurological disorder; Huntington's disease; Alzheimer's disease;

KW Parkinson's disease; prion disease; frontotemporal dementia;

KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;

KW dentatorubal-pallidoluysian atrophy; spinocerebellar ataxia type 1; SCA2;

KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.

XX

OS Synthetic.

XX

PN WO200106989-A2.

KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
 KW polyglutamine expansion; Huntington's disease; Alzheimer's disease;
 KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
 KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
 KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1
 FT /note= "GST protein connected to the N-terminal"
 FT Misc-difference 25
 FT /note= "polyglutamine expansion that can comprise upto 51
 FT glutamines"
 XX
 PN WO9906838-A2.
 XX
 PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98WO-EP004810.
 XX
 PR 01-AUG-1997; 97EP-00113320.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Wanker E, Lehrach H, Scherzinger E, Bates G;
 XX
 DR WPI; 1999-153955/13.
 XX
 PT Detecting amyloid-like fibrils or protein aggregates insoluble in
 PT detergent or urea - from their retention on a filter, used for diagnosis,
 PT particularly of diseases associated with polyglutamine expansion.
 XX
 PS Example 1; Fig 2; 56pp; English.
 XX
 CC The invention relates to the detection of amyloid-like fibrils or protein
 CC aggregates, insoluble in detergents or urea. The method comprises: (a)
 CC applying material suspected of containing protein aggregates to a filter;
 CC and (b) detecting retention of protein aggregates on the filter. This
 CC method also helps to identify inhibitors of protein aggregates formation.
 CC The method is particularly used to detect protein aggregates that are
 CC indicative of disease, for assessing onset or progression of the
 CC diseases. The inhibitors identified are potential therapeutic agents for
 CC treating the diseases. Other applications include detection of inclusion
 CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
 CC associated with polyglutamine expansion are particularly diagnosed, e.g.
 CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
 CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
 CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
 CC scrapie. The protein aggregates can now be detected simply, routinely and
 CC rapidly, without requiring sophisticated equipment. The method can be
 CC made quantitative, by analysing a series of dilutions, and can be
 CC automated to allow many samples to be analysed on the same filter. The
 CC present sequence represents the Huntington's gene exon 1 translation
 CC product which is connected to a GST protein to form a fusion protein. The
 CC sequence of the GST protein is not indicated

SQ Sequence 108 AA;

Qy 7 MATLEKLMKAFESLKS FQQ 58
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 MATLEKLMKAFESLKS FQQ 59

AAW95076

PS Example 1; Fig 2; 62pp; English.

The invention relates to a composition comprising a fusion protein of (i) (poly)peptide that increases solubility and/or prevents aggregation of fusion protein, and (ii) amyloidogenic (poly)peptide that can self-assemble into amyloid-like fibrils or protein aggregates. Host cells transformed with a vector containing the nucleic acid encoding the fusion protein are used for the recombinant expression of the fusion protein. The composition is used to detect onset and progression of diseases associated with fibrils/protein aggregates. It is potentially useful for treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-repeat expansion conditions such as Huntington's disease (HD), spinal and bulbar muscular atrophy, dentatorubral pallidoluysian atrophy, spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on release of the amyloidogenic polypeptide from fusion protein have a precise starting time for aggregate formation, allowing kinetic measurements, and use of an enzyme for cleavage allows testing under physiological conditions. The present sequence represents the Huntington's gene exon 1 translation product which is connected to a GST protein to form a fusion protein. The sequence of the GST protein is not indicated

SQ Sequence 108 AA;

Query Match 68.5%; Score 255; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

Search completed: March 12, 2004, 15:38:31
Job time : 54.6471 secs

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:38:34 ; Search time 15.6471 Seconds
 (without alignments)
 250.755 Million cell updates/sec

Title: US-09-620-955B-11
 Perfect score: 372
 Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 76

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Issued_Patents_AA:*
 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query					Description
	Score	Match Length	DB	ID		
1	207	55.6	910	4	US-08-997-685A-2	Sequence 2, Appli
2	204	54.8	513	3	US-09-041-886-28	Sequence 28, Appl
3	204	54.8	530	3	US-09-041-886-29	Sequence 29, Appl
4	204	54.8	552	3	US-09-041-886-30	Sequence 30, Appl
5	204	54.8	589	3	US-09-041-886-31	Sequence 31, Appl
6	204	54.8	2074	4	US-09-491-356C-9	Sequence 9, Appli
7	204	54.8	3144	1	US-08-246-982A-6	Sequence 6, Appli
8	204	54.8	3144	1	US-08-453-265-6	Sequence 6, Appli
9	204	54.8	3144	2	US-08-457-273B-42	Sequence 42, Appl
10	204	54.8	3144	3	US-08-556-419-21	Sequence 21, Appl
11	204	54.8	3144	3	US-09-041-886-15	Sequence 15, Appl

12	199.5	53.6	2023	4	US-09-491-356C-8	Sequence 8, Appli
13	170	45.7	1420	4	US-09-125-635-4	Sequence 4, Appli
14	166.5	44.8	678	5	PCT-US93-03027-3	Sequence 3, Appli
15	162.5	43.7	788	2	US-08-918-914-4	Sequence 4, Appli
16	162	43.5	1282	4	US-09-543-681A-5419	Sequence 5419, Ap
17	160	43.0	2703	1	US-08-185-432-19	Sequence 19, Appl
18	160	43.0	2703	4	US-08-899-232-4	Sequence 4, Appli
19	159	42.7	816	2	US-08-267-803B-9	Sequence 9, Appli
20	159	42.7	816	3	US-09-041-886-17	Sequence 17, Appl
21	157	42.2	528	4	US-09-086-663A-82	Sequence 82, Appl
22	157	42.2	548	4	US-09-086-663A-71	Sequence 71, Appl
23	157	42.2	596	4	US-09-086-663A-2	Sequence 2, Appli
24	157	42.2	596	4	US-09-086-663A-80	Sequence 80, Appl
25	156	41.9	729	4	US-09-625-188-20	Sequence 20, Appl
26	155	41.7	1162	2	US-08-728-323A-2	Sequence 2, Appli
27	155	41.7	1162	4	US-09-298-568-2	Sequence 2, Appli
28	155	41.7	1162	4	US-09-410-399-2	Sequence 2, Appli
29	154.5	41.5	903	2	US-08-853-310-2	Sequence 2, Appli
30	154	41.4	428	1	US-08-190-802A-29	Sequence 29, Appl
31	154	41.4	428	3	US-08-477-346-29	Sequence 29, Appl
32	154	41.4	428	4	US-08-473-089-29	Sequence 29, Appl
33	154	41.4	428	4	US-08-487-072A-29	Sequence 29, Appl
34	153	41.1	1003	4	US-09-521-511C-11	Sequence 11, Appl
35	153	41.1	1507	4	US-09-914-259-37	Sequence 37, Appl
36	152.5	41.0	1402	4	US-09-125-635-12	Sequence 12, Appl
37	152	40.9	513	3	US-09-100-193-3	Sequence 3, Appli
38	151	40.6	542	1	US-07-814-964-13	Sequence 13, Appl
39	151	40.6	542	1	US-08-258-442-13	Sequence 13, Appl
40	151	40.6	542	1	US-08-328-809-8	Sequence 8, Appli
41	151	40.6	542	4	US-08-866-840-8	Sequence 8, Appli
42	151	40.6	542	5	PCT-US92-11107-13	Sequence 13, Appl
43	150.5	40.5	885	4	US-09-914-259-10	Sequence 10, Appl
44	148.5	39.9	361	1	US-08-415-751-4	Sequence 4, Appli
45	148	39.8	360	2	US-08-531-927B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-997-685A-2

; Sequence 2, Application US/08997685A

; Patent No. 6551821

; GENERAL INFORMATION:

; APPLICANT: The Trustees of Columbia University

; APPLICANT: Kandel, Eric

; TITLE OF INVENTION: Brain Cyclic Nucleotide Gated Ion Channel and Uses
Thereof

; FILE REFERENCE: 0575/54806

; CURRENT APPLICATION NUMBER: US/08/997,685A

; CURRENT FILING DATE: 1997-12-12

; NUMBER OF SEQ ID NOS: 60

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 910

; TYPE: PRT

; ORGANISM: mouse

```

; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (130)..(148)
; OTHER INFORMATION: S1
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (164)..(185)
; OTHER INFORMATION: S2
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (208)..(229)
; OTHER INFORMATION: S3
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (243)..(271)
; OTHER INFORMATION: S4
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (291)..(313)
; OTHER INFORMATION: S5
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (332)..(358)
; OTHER INFORMATION: P
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (367)..(387)
; OTHER INFORMATION: S6
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (472)..(602)
; OTHER INFORMATION: CNB
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AAC53518
; DATABASE ENTRY DATE: 1997-12-27
; RELEVANT RESIDUES: (1)..(910)
US-08-997-685A-2

```

```

Query Match          55.6%; Score 207; DB 4; Length 910;
Best Local Similarity 87.5%; Pred. No. 3.3e-16;
Matches 42; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      24 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLQPGST 71
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      735 QTQTQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQPQTPGSS 782

```

```

RESULT 2
US-09-041-886-28
; Sequence 28, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72

```

```

;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Campbell & Flores LLP
;   STREET:  4370 La Jolla Village Drive, Suite 700
;   CITY:  San Diego
;   STATE:  California
;   COUNTRY:  United States
;   ZIP:  92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/09/041,886
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Campbell, Cathryn A.
;   REGISTRATION NUMBER:  31,815
;   REFERENCE/DOCKET NUMBER:  P-LJ 2626
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (619) 535-9001
;   TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  28:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  513 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
US-09-041-886-28

```

```

Query Match          54.8%;  Score 204;  DB 3;  Length 513;
Best Local Similarity 72.6%;  Pred. No. 4e-16;
Matches  45;  Conservative  0;  Mismatches  17;  Indels  0;  Gaps  0;

```

```

QY      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
          |||||||||||||||||||||||||||||||||||||||
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPQA 60

QY      67 QP 68
          ||
Db      61 QP 62

```

RESULT 3

US-09-041-886-29

```

; Sequence 29, Application US/09041886
; Patent No. 6235872
;   GENERAL INFORMATION:
;   APPLICANT:  Bredeesen, Dale E.
;   APPLICANT:  Rabizadeh, Sharroz
;   TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;   TITLE OF INVENTION:  Polypeptides and Methods of Use
;   NUMBER OF SEQUENCES:  72
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Campbell & Flores LLP
;   STREET:  4370 La Jolla Village Drive, Suite 700

```

```

; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-041-886-29

```

```

Query Match          54.8%; Score 204; DB 3; Length 530;
Best Local Similarity 72.6%; Pred. No. 4.1e-16;
Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

```

Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
          |||||||||||||||||||||||||||||||||||||||
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPQA 60

Qy      67 QP 68
          ||
Db      61 QP 62

```

RESULT 4

US-09-041-886-30

```

; Sequence 30, Application US/09041886
; Patent No. 6235872

```

; GENERAL INFORMATION:

```

; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States

```

```

;      ZIP: 92122
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Campbell, Cathryn A.
;      REGISTRATION NUMBER: 31,815
;      REFERENCE/DOCKET NUMBER: P-LJ 2626
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (619) 535-9001
;      TELEFAX: (619) 535-8949
;      INFORMATION FOR SEQ ID NO: 30:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 552 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-09-041-886-30

```

```

Query Match          54.8%; Score 204; DB 3; Length 552;
Best Local Similarity 72.6%; Pred. No. 4.3e-16;
Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

```

Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
      ||||||||||||||||||||||||||||||||||| | | |
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPPQLPQPPQA 60

Qy      67 QP 68
      ||
Db      61 QP 62

```

RESULT 5

US-09-041-886-31

```

; Sequence 31, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;      INFORMATION FOR SEQ ID NO:  31:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  589 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  peptide
US-09-041-886-31

```

```

Query Match          54.8%;  Score 204;  DB 3;  Length 589;
Best Local Similarity 72.6%;  Pred. No. 4.6e-16;
Matches  45;  Conservative  0;  Mismatches  17;  Indels  0;  Gaps  0;

```

```

Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
          |||||||||||||||||||||||||||||||||||  |  |  |
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPPQLPQPPPPQA 60

Qy      67 QP 68
          ||
Db      61 QP 62

```

RESULT 6

US-09-491-356C-9

```

; Sequence 9, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; APPLICANT: Delisi, Lynn
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF
XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT APPLICATION NUMBER: US/09/491,356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 2074
; TYPE: PRT

```


; ORGANISM: Mus musculus
US-09-491-356C-9

Query Match 54.8%; Score 204; DB 4; Length 2074;
Best Local Similarity 67.2%; Pred. No. 1.8e-15;
Matches 45; Conservative 2; Mismatches 6; Indels 14; Gaps 1;

QY 24 QQQQQQQQQQQQQQQQQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQLQPG 69
||||| |
Db 1956 QQQQQQQQQQQQQQQQQQYHIRQQQQQQQMLRQQQQQQQQQQQQQQQQQQQQQQQPH 2015
QY 70 STRAAAS 76
: |:
Db 2016 QQQQQQAA 2022

RESULT 7

US-08-246-982A-6

; Sequence 6, Application US/08246982A
; Patent No. 5686288

; GENERAL INFORMATION:

; APPLICANT: MacDonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/246,982A
; FILING DATE: May 20, 1994
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldstein, Jorge, A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3880002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-246-982A-6

Query Match 54.8%; Score 204; DB 1; Length 3144;
Best Local Similarity 72.6%; Pred. No. 2.8e-15;
Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
      |||||||||||||||||||||||||||||||||||  |  |  |
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPPQA 60

Qy      67 QP 68
      ||
Db      61 OP 62

```

RESULT 8

US-08-453-265-6

; Sequence 6, Application US/08453265

; Patent No. 5693757

; GENERAL INFORMATION:

; APPLICANT: MacDonald, Marcy E.

APPLICANT: Ambrose, Christine M.

APPLICANT: Duyao, Mabel P.

APPLICANT: Gusella, James F.

10 TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue

; CITY: Washington

STATE: D.C.

; COUNTRY: U.S.A.

ZIP: 20005

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/453,265

FILING DATE: 30-MAY-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, Steven R.

REGISTRATION NUMBER: 36,203

REFERENCE/DOCKET NUMBER: 0609.3880003

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

LENGTH: 3144 amino acids

```

;      TYPE:  amino acid

```

```
; TOPOLOGY: linear
```

; MOLECULE TYPE: protein

US-08-453-265-6

Query Match 54.8%; Score 204; DB 1; Length 3144;

Best Local Similarity 72.6%; Pred. No. 2.8e-15;
Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
          |||||||||||||||||||||||||||||||||||
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPPQA 60

Qy      67 QP 68
          ||
Db      61 QP 62
```

RESULT 9

US-08-457-273B-42
; Sequence 42, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-457-273B-42

Query Match 54.8%; Score 204; DB 2; Length 3144;
Best Local Similarity 72.6%; Pred. No. 2.8e-15;

Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQQQL 66
 |||
 Db 1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPPQA 60

Qy 67 QP 68
 ||
 Db 61 QP 62

RESULT 10

US-08-556-419-21
 ; Sequence 21, Application US/08556419C
 ; Patent No. 6093549
 ; GENERAL INFORMATION:
 ; APPLICANT: Ross, Christopher
 ; APPLICANT: Li, Xiao-Jiang
 ; APPLICANT: Li, Shi-Hua
 ; APPLICANT: Sharp, Alan
 ; APPLICANT: Lanahan, Anthony
 ; APPLICANT: Worley, Paul
 ; APPLICANT: Snyder, Solomon
 ; TITLE OF INVENTION: Huntingtin-associated protein
 ; FILE REFERENCE: 01107.52271
 ; CURRENT APPLICATION NUMBER: US/08/556,419C
 ; CURRENT FILING DATE: 1995-11-09
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 21
 ; LENGTH: 3144
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-08-556-419-21

Query Match 54.8%; Score 204; DB 3; Length 3144;
 Best Local Similarity 72.6%; Pred. No. 2.8e-15;
 Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQQQL 66
 |||
 Db 1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPPQA 60

Qy 67 QP 68
 ||
 Db 61 QP 62

RESULT 11

US-09-041-886-15
 ; Sequence 15, Application US/09041886
 ; Patent No. 6235872
 ; GENERAL INFORMATION:
 ; APPLICANT: Bredesen, Dale E.
 ; APPLICANT: Rabizadeh, Sharroz
 ; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
 ; TITLE OF INVENTION: Polypeptides and Methods of Use

```

;   NUMBER OF SEQUENCES:  72
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Campbell & Flores LLP
;   STREET:  4370 La Jolla Village Drive, Suite 700
;   CITY:  San Diego
;   STATE:  California
;   COUNTRY:  United States
;   ZIP:  92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/09/041,886
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Campbell, Cathryn A.
;   REGISTRATION NUMBER:  31,815
;   REFERENCE/DOCKET NUMBER:  P-LJ 2626
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (619) 535-9001
;   TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  15:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  3144 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-09-041-886-15

```

```

Query Match          54.8%;  Score 204;  DB 3;  Length 3144;
Best Local Similarity 72.6%;  Pred. No. 2.8e-15;
Matches  45;  Conservative  0;  Mismatches  17;  Indels  0;  Gaps  0;

```

```

Qy      7 MATLEKIMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66 /
      |||||||||||||||||||||||||||||||||||||||||||||  |  |  |
Db      1 MATLEKIMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPPQLPQPPPPQA 60

Qy      67 QP 68
      ||
Db      61 QP 62

```

```

RESULT 12
US-09-491-356C-8
; Sequence 8, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; APPLICANT: Delisi, Lynn
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF
XQ13
; FILE REFERENCE: 9465.6USI1
; CURRENT APPLICATION NUMBER: US/09/491,356C

```

```
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 2023
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-491-356C-8
```

```
Query Match          53.6%; Score 199.5; DB 4; Length 2023;
Best Local Similarity 57.3%; Pred. No. 5.7e-15;
Matches 47; Conservative 5; Mismatches 17; Indels 13; Gaps 1;
```

```
Qy      6 SMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQ-----QQQQQQQQQQQQQQ 52
      | | | : : : | | | | | | | | | | | | | | | | | | | | | | | |
Db      1891 STAILPEQQQQQQQQQQQQQQQQQQQQQQQQQQQQYHIRQQQQQQILRQQQQQQQQQQQQQQ 1950

Qy      53 QQQQQQQQQQQQQQLQPGSTRAA 74
      | | | | | | | | | | | : | |
Db      1951 QQQQQQQQQQQQHQQQQQQQQAA 1972
```

RESULT 13

US-09-125-635-4

```
; Sequence 4, Application US/09125635
; Patent No. 6562589
; GENERAL INFORMATION:
; APPLICANT: THE UNITED STATES OF AMERICA represented by THE SE
; TITLE OF INVENTION: AIB1, A novel steroid receptor co-activator
; FILE REFERENCE: 49944
; CURRENT APPLICATION NUMBER: US/09/125,635
; CURRENT FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: 60/049,728
; PRIOR FILING DATE: 1997-06-17
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-125-635-4
```

```
Query Match          45.7%; Score 170; DB 4; Length 1420;
Best Local Similarity 66.7%; Pred. No. 1.1e-11;
Matches 36; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
```

```
Qy      23 FQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLQPGSTRAAAS 76
      | : | | : | | | | | | | | | | | | | | | | | | | | | | |
Db      1234 FRQQRVAMMQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQTQAFSPPPNVTASPS 1287
```

RESULT 14

PCT-US93-03027-3

; Sequence 3, Application PC/TUS9303027
; GENERAL INFORMATION:
; APPLICANT: LEONARD, WARREN; TOLEDANO,
; APPLICANT: MICHEL
; TITLE OF INVENTION: CONTROL AND/OR
; TITLE OF INVENTION: PREVENTION OF BINDING OF NF- B/REL/DORSAL
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03027
; FILING DATE: 19930401
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,987
; FILING DATE: 06-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: P-36,434
; REFERENCE/DOCKET NUMBER: 2026-4010 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 678
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; STRAIN: Oregon R
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: embryo
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; FEATURE:
; NAME/KEY: Dorsal protein
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: D.melanogaster
; OTHER INFORMATION: embryonic polarity (dorsal) protein

```

;   OTHER INFORMATION:  containing region of high similarity
;   OTHER INFORMATION:  with proteins of Rel family.
;   PUBLICATION INFORMATION:
;   AUTHORS:  Steward, R.
;   TITLE:  Dorsal, an embryonic polarity
;   TITLE:  gene in Drosophila, is homologous to
;   TITLE:  the vertebrate proto-oncogene, c-rel.
;   JOURNAL:  Science
;   VOLUME:  238
;   ISSUE:
;   PAGES:  692-694
;   DATE:  1987
;   DOCUMENT NUMBER:
;   FILING DATE:
;   PUBLICATION DATE:
;   RELEVANT RESIDUES IN SEQ ID NO:
PCT-US93-03027-3

```

```

Query Match          44.8%;  Score 166.5;  DB 5;  Length 678;
Best Local Similarity 58.7%;  Pred. No. 1.2e-11;
Matches  37;  Conservative  1;  Mismatches  6;  Indels  19;  Gaps  1;

```

```

Qy      24 QQQQQQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 64
          |||||          ||| ||||| | | ||| |||||:|
Db      448 QQQQQQQYGGATDLGSNNPFAQQVLAQQQQHQQQQQHQHQHQQQQQQQQQQQEQQ 507

Qy      65 QLQ 67
          ||
Db      508 SLQ 510

```

RESULT 15

US-08-918-914-4

```

; Sequence 4, Application US/08918914
; Patent No. 5876963

```

; GENERAL INFORMATION:

```

;   APPLICANT:  Mitchell, Peter
;   APPLICANT:  Hutchinson, Nancy
;   APPLICANT:  Lawton, Michael
;   APPLICANT:  Magna, Holly
;   APPLICANT:  Yocum, Sue
;   APPLICANT:  Murry, Lynn E.

```

; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

```

;   ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;   STREET:  3174 Porter Dr.
;   CITY:  Palo Alto
;   STATE:  CA
;   COUNTRY:  USA
;   ZIP:  94304

```

; COMPUTER READABLE FORM:

```

;   MEDIUM TYPE:  Diskette
;   COMPUTER:  IBM Compatible
;   OPERATING SYSTEM:  DOS
;   SOFTWARE:  FastSEQ for Windows Version 2.0

```

; CURRENT APPLICATION DATA:


```

; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1070094
US-08-918-914-4

```

```

Query Match          43.7%; Score 162.5; DB 2; Length 788;
Best Local Similarity 46.8%; Pred. No. 4.3e-11;
Matches 37; Conservative 5; Mismatches 6; Indels 31; Gaps 1;

```

```

Qy      24 QQQQQQQQQQQQQQQQQQ-----QQQQQQQQQQQQQ 52
      || ||||| ||||| |||||: ||||| |||||
Db      244 QQHQQQQQQQQQQQQQQQRPPQPQPQPQPQPQRPPQPQSFSGTHELHLQRQREQQQQQQQQ 303
Qy      53 QQQQQQQQQQQQQQLQPGST 71
      ||||| ||||| || || ||
Db      304 QQQQQQQRQQNPQQQPQQT 322

```

```

Search completed: March 12, 2004, 15:42:41
Job time : 16.6471 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:36:59 ; Search time 12.6667 Seconds
(without alignments)
577.149 Million cell updates/sec

Title: US-09-620-955B-11
Perfect score: 372
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 76

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*

- 1: pirl:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		% Query				ID	Description
No.	Score	Match	Length	DB			
1	224	60.2	646	2	D82493		conserved hypothet
2	215.5	57.9	1969	2	T08875		histidine kinase h
3	215	57.8	1457	2	T14577		protein kinase Yak
4	213	57.3	905	1	RGBYS5		regulatory protein
5	211.5	56.9	339	1	TWHU2D		transcription init
6	208	55.9	1905	2	T18267		multidrug resistan
7	204.5	55.0	1761	2	T13675		hypothetical prote
8	204	54.8	139	2	A26892		Mopa box protein -
9	204	54.8	3144	2	A46068		Huntington disease
10	198.5	53.4	1023	2	T13068		CLOCK protein - fr
11	198.5	53.4	1027	2	T13071		CLOCK protein - fr
12	197.5	53.1	4957	2	T03455		ALR protein - huma
13	197.5	53.1	5262	2	T03454		ALR protein - huma

14	191	51.3	966	2	S25365	CYC8 protein - yea
15	190	51.1	758	2	S54522	hypothetical prote
16	188.5	50.7	930	2	T08588	hypothetical prote
17	184	49.5	648	1	JQ1150	protein kinase (EC
18	182.5	49.1	1225	2	T09057	probable protein-h
19	182	48.9	429	2	S29565	apolipoprotein A-I
20	182	48.9	1154	2	S69206	regulator protein
21	181	48.7	873	2	B53225	ecdysone-induced p
22	181	48.7	2150	2	S71629	sensory transducti
23	180.5	48.5	613	2	S27770	hypothetical prote
24	180	48.4	1390	2	T14004	trfA protein - sli
25	179	48.1	1010	2	T13167	Lola-like protein
26	178.5	48.0	1596	2	A33106	neurogenic locus m
27	176	47.3	1015	2	T13062	CLOCK protein - fr
28	175	47.0	3190	2	T13828	CREB-binding prote
29	174.5	46.9	354	2	S65687	(A+T)-stretch-bind
30	174	46.8	770	2	T51024	related to C2H2 zi
31	173.5	46.6	398	2	S21883	bZIP transcription
32	173.5	46.6	521	2	A40252	elav protein - fru
33	172.5	46.4	1655	2	T13998	gene mastermind pr
34	172	46.2	534	2	G86385	hypothetical prote
35	172	46.2	756	2	S60966	probable protein k
36	172	46.2	1090	2	A41696	regulatory protein
37	171	46.0	1572	2	S45251	SNF2alpha protein
38	170	45.7	1081	2	S66736	transcription acti
39	170	45.7	1424	2	T03851	thyroid hormone re
40	169.5	45.6	678	1	A30350	dorsal protein - f
41	169	45.4	2038	2	A43742	female sterile hom
42	169	45.4	3848	2	T17414	TipC protein - sli
43	168.5	45.3	473	2	T09580	hypothetical prote
44	168	45.2	539	2	S57972	hypothetical prote
45	168	45.2	1403	2	S24548	homeotic protein p

ALIGNMENTS

RESULT 1

D82493

conserved hypothetical protein VCA0171 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C;Species: *Vibrio cholerae*

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C;Accession: D82493

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill, S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.; Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.; Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: D82493

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-646 <HEI>
A;Cross-references: GB:AE004357; GB:AE003853; NID:g9657547; PIDN:AAF96084.1;
GSPDB:GN00127; TIGR:VCA0171
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0171
A;Map position: 2

Query Match 60.2%; Score 224; DB 2; Length 646;
Best Local Similarity 80.7%; Pred. No. 7.4e-12;
Matches 46; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 20 LKSFQQQLPGSTRAAAS 76
:|: ||||| : : ||
Db 440 VKAAQQQDSSSGAS 496

RESULT 2

T08875

histidine kinase homolog DHKB - slime mold (Dictyostelium discoideum)

N;Alternate names: hybrid histidine kinase DHKB

C;Species: Dictyostelium discoideum

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-May-2000

C;Accession: T08875

R;Zinda, M.J.; Singleton, C.K.

Dev. Biol. 196, 171-183, 1998

A;Title: The hybrid histidine kinase dhkB regulates spore germination in Dictyostelium discoideum.

A;Reference number: Z16506; MUID:98248997; PMID:9576830

A;Accession: T08875

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1969 <SIN>

A;Cross-references: EMBL:AF024654; NID:g2460282; PID:g2460283

A;Experimental source: strain KAx3

C;Genetics:

A;Gene: dhkB

A;Introns: 790/3

C;Superfamily: response regulator homology

C;Keywords: protein kinase; transmembrane protein

F;1841-1964/Domain: response regulator homology <RRH>

Query Match 57.9%; Score 215.5; DB 2; Length 1969;
Best Local Similarity 51.5%; Pred. No. 9.5e-11;
Matches 51; Conservative 9; Mismatches 14; Indels 25; Gaps 2;

Qy 2 VPRGSMATLEKLMKAFESLKSF-----QQQQQQQQQQQQQQQ 37
| ||:::| |:: | |:| |||||
Db 1675 VRSGSLSSL-KPLREDEELESISDDHTSHLKGSSHSINQQIPSTIQQQQQQQQQQQQQQ 1733

Qy 38 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLPGSTRAAAS 76
|||||||: ||| :| :| |
Db 1734 QQQQQQQQQQQQQQQQQQQQQQQQKPQQQQKPTTTTTTTS 1772

RESULT 3

T14577

protein kinase YakA (EC 2.7.1.-) - slime mold (*Dictyostelium discoideum*)
 C;Species: *Dictyostelium discoideum*
 C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C;Accession: T14577
 R;Kuspa, A.; Lu, S.; Souza, G.M.
 submitted to the EMBL Data Library, January 1998
 A;Description: YakA, a protein kinase required for the growth to development transition in *Dictyostelium*.
 A;Reference number: Z18146
 A;Accession: T14577
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-1457 <KUS>
 A;Cross-references: EMBL:AF045453; NID:g2854116; PID:g2854117; PIDN:AAC02554.1
 C;Genetics:
 A;Gene: yakA
 C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protein kinase

Query Match 57.8%; Score 215; DB 2; Length 1457;
 Best Local Similarity 66.2%; Pred. No. 8.1e-11;
 Matches 45; Conservative 7; Mismatches 12; Indels 4; Gaps 1;

```

QY      2 VPRGSMATLEKLMKAFESLKSFOQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ----QQQQ 57
      :|: ||   ::   : :  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      575 IPQHSMNLGNQILNQHLFQQLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHNFQQQQ 634

QY      58 QQQQQQQQ 65
      ||| | | | |
Db      635 QQQQQQQQ 642
  
```

RESULT 4

RGBYS5

regulatory protein SNF5 - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: protein YBR2036; protein YBR289w

C;Species: *Saccharomyces cerevisiae*

C;Date: 30-Sep-1991 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000

C;Accession: S44551; S46171; A36375; S12067; S39145

R;Holmstrom, K.; Brandt, T.; Kallesoe, T.

Yeast 10(Suppl.A), S47-S62, 1994

A;Title: The sequence of a 32420 bp segment located on the right arm of chromosome II from *Saccharomyces cerevisiae*.

A;Reference number: S44537; MUID:94378722; PMID:8091861

A;Accession: S44551

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-905 <HOL>

A;Cross-references: EMBL:X76053; NID:g600025; PIDN:CAA53652.1; PID:g429134

R;Brandt, T.; Christiansen, C.; Holmstroem, K.; Kallesoe, T.

submitted to the Protein Sequence Database, August 1994

A;Reference number: S46157

A;Accession: S46171

A;Molecule type: DNA

A;Residues: 1-905 <BRA>

A;Cross-references: EMBL:Z36158; NID:g536741; PIDN:CAA85254.1; PID:g536742;

GSPDB:GN00002; MIPS:YBR289w

A;Title: Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID).

A;Reference number: S10944; MUID:90326195; PMID:2374612

A;Accession: S10944

A;Molecule type: mRNA

A;Residues: 1-91,96-339 <HOF>

A;Cross-references: EMBL:X54993; NID:g37065; PIDN:CAA38736.1; PID:g37066

R;Kao, C.; Lieberman, P.; Schmidt, M.; Zhou, Q.; Pei, R.; Berk, A.J.

Science 248, 1626, 1990

A;Title: Cloning of the human TATA binding factor: Expression of a transcriptionally active TFIID protein.

A;Reference number: I60128

A;Accession: I60128

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-186,'R',188-299,'MIKPR',300-339 <RES>

A;Cross-references: GB:M34960; NID:g339493; PID:g339494

C;Genetics:

A;Gene: GDB:TBP; GTF2D1

A;Cross-references: GDB:138768; OMIM:600075

A;Map position: 6q27-6q27

C;Superfamily: human transcription initiation factor IID

C;Keywords: alternative splicing; DNA binding; nucleus; transcription initiation

F;55-95/Region: glutamine-rich

Query Match 56.9%; Score 211.5; DB 1; Length 339;

Best Local Similarity 67.7%; Pred. No. 4.8e-11;

Matches 44; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

```
Qy      1 LVPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60
          ::| |: | : : || ::||:|||||
Db      31 MMPYGTGLTPQPIQNT-NSLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 89

Qy      61 QQQQQ 65
          |||||
Db      90 QQQQQ 94
```

RESULT 6

T18267

multidrug resistance protein - slime mold (Dictyostelium discoideum)

C;Species: Dictyostelium discoideum

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T18267

R;Shaulsky, G.; Kuspa, A.; Loomis, W.F.

submitted to the EMBL Data Library, January 1995

A;Description: An MDR transporter/serine protease gene is required for prestalk specialization in Dictyostelium.

A;Reference number: Z18850

A;Accession: T18267

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1905 <SHA>

A;Cross-references: EMBL:U20432; NID:g664839; PID:g664840; PIDN:AAA62212.1

C;Genetics:

A;Gene: tagB

Query Match 55.9%; Score 208; DB 2; Length 1905;
Best Local Similarity 82.4%; Pred. No. 3.9e-10;
Matches 42; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

RESULT 7

hypothetical protein EG0002.3 - fruit fly (*Drosophila melanogaster*)
C;Species: *Drosophila melanogaster*
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13675
R;Bolshakov, V.; Borkova, D.; Minana, B.; Kafatos, F.
submitted to the EMBL Data Library, September 1998
A;Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.
A;Reference number: Z17698
A;Accession: T13675
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1761 <BOL>
A;Cross-references: EMBL:AL031130; NID:e1316407; PID:e1316410; PIDN:CAA20016.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025376
A;Introns: 143/3; 237/3; 280/3
A;Note: EG:EG0002.3

```

Qy          3 PRGSMATLEKLMKAFESLKSFOQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 62
              | : | :: : : | | | | | | | | | | | | | | | | | | | | | | | : :
Db        1474 PAGATADMQRVQRMQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQLYR 1528

Qy          63 QQQLQPGSTRAAAS 76
              ||| | |
Db        1529 QQQLHRTSMGSIAS 1542

```

A26892

C;Species: Mus musculus (house mouse)

R;Duboule, D.; Haenlin, M.; Galliot, B.; Mohier, E.

A;Title: DNA sequences homologous to the Drosophila opa repeat are present in murine mRNAs that are differentially expressed in fetuses and adult tissues.

A;Accession: A26892

A;Residues: 1-139 <DUB>

Query Match 54.8%; Score 204; DB 2; Length 139;
Best Local Similarity 67.2%; Pred. No. 9.7e-11;
Matches 45; Conservative 2; Mismatches 6; Indels 14; Gaps 1;

```
Qy      24  QQQQQQQQQQQQQQQQQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQLQPG 69
          |||
Db      21  QQQQQQQQQQQQQQQQQQYHIRQQQQQQQMLRQQQQQQQQQQQQQQQQQQQQQQQPH 80
          |||

Qy      70  STRAAAS 76
          : |:
Db      81  QQQQQQAA 87
```

RESULT 9

A46068

Huntington disease-associated protein - human

C;Species: Homo sapiens (man)

C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999

C;Accession: A46068; I54337

R;MacDonald, M.E.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.; Groot, N.; MacFarlane, H.; Jenkins, B.; Anderson, M.A.; Wexler, N.S.; Gusella, J.F.; Bates, G.P.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott, R.; Zehetner, G.; Sedlacek, Z.; Poustka, A.; Frischauf, A.M.; Buckler, A.J.; Church, D.; Doucette-Stamm, L.; O'Donovan, M.C.; Riba-Ramirez, L.; Shah, M.; Stanton, V.P.; Strobel, S.A.; Draths, K.M.

Cell 72, 971-983, 1993

A;Authors: Wales, J.L.; Dervan, P.; Housman, D.E.; Altherr, M.; Shiang, R.; Thompson, L.; Fielder, T.; Wasmuth, J.J.; Tagle, D.; Valdes, J.; Elmer, L.; Allard, M.; Castilla, L.; Swaroop, M.; Blanchard, K.; Collins, F.S.; Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harper, P.S.

A;Title: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.

A;Reference number: A46068; MUID:93208892; PMID:8458085

A;Accession: A46068

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-3144 <MAC>

A;Cross-references: GB:L12392

R;Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.; MacDonald, H.; Nasir, J.; Delaney, A.; Goldberg, Y.P.; Hayden, M.R.

Hum. Mol. Genet. 2, 1541-1545, 1993

A;Title: Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression.

A;Reference number: I54337; MUID:94093536; PMID:7903579

A;Accession: I54337

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 2563-3144 <RES>

A;Cross-references: GB:L20431; NID:g398028; PIDN:AAA52702.1; PID:g398029

C;Genetics:

A;Gene: GDB:HD

A;Cross-references: GDB:119307; OMIM:143100

A;Map position: 4p16.3-4p16.3

Query Match 54.8%; Score 204; DB 2; Length 3144;

Best Local Similarity 72.6%; Pred. No. 1.3e-09;
Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
          ||||||||||||||||||||||||||||||||||||||| | | |
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPPQA 60

Qy      67 QP 68
          ||
Db      61 QP 62
```

RESULT 10

T13068

CLOCK protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C;Accession: T13068

R;Darlington, T.K.; Wager-Smith, K.; Ceriani, M.F.; Staknis, D.; Gekakis, N.;
Steeves, T.D.L.; Weitz, C.J.; Takahashi, J.S.; Kay, S.A.
Science 280, 1599-1603, 1998

A;Title: Closing the circadian loop: CLOCK-induced transcription of its own
inhibitors per and tim.

A;Reference number: Z17599; MUID:98279147; PMID:9616122

A;Accession: T13068

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1023 <DAR>

A;Cross-references: EMBL:AF067207; NID:g3192866; PID:g3192867; PIDN:AAD10630.1

C;Genetics:

A;Cross-references: FlyBase:FBgn0023076

C;Function:

A;Description: required for circadian behavioral rhythms

Query Match 53.4%; Score 198.5; DB 2; Length 1023;
Best Local Similarity 76.8%; Pred. No. 1.5e-09;
Matches 43; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

```
Qy      13 LMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLQ 67
          | : :| || ||| ||||||||||||||||||||||||||||||| |||
Db      779 LQQQHQSLSQLQQHTQQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLQLQ 834
```

RESULT 11

T13071

CLOCK protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C;Accession: T13071

R;Bae, K.; Lee, C.; Sidote, D.; Chuang, K.Y.; Ederly, I.
Mol. Cell. Biol. 18, 6142-6151, 1998

A;Title: Circadian regulation of a drosophila homolog of the mammalian clock
gene: PER and TIM function as positive regulators.

A;Reference number: Z17601; MUID:98414630; PMID:9742131

A;Accession: T13071

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

C;Accession: T03454
 R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, T.; Alder, H.; Croce, C.M.; Huebner, K.; Mazo, A.; Canaani, E. Oncogene 15, 549-560, 1997
 A;Title: Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia, and to Drosophila trithorax.
 A;Reference number: Z14954; MUID:97388474; PMID:9247308
 A;Accession: T03454
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-5262 <PRA>
 A;Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285
 C;Genetics:
 A;Gene: ALR
 A;Map position: 12
 C;Superfamily: acute lymphoblastic leukemia protein, ALR type
 C;Keywords: alternative splicing

Query Match 53.1%; Score 197.5; DB 2; Length 5262;
 Best Local Similarity 66.2%; Pred. No. 6.8e-09;
 Matches 43; Conservative 6; Mismatches 15; Indels 1; Gaps 1;

Qy 3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 62
 | ||: |:: : : : ||||| |||| | ||||| ||||| ||||| ||||| ||||| |||||
 Db 3618 PMGSLQQLQQ-QQQLQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQQQLQQQQQLQQQ 3676

Qy 63 QQQLQ 67
 |||||
 Db 3677 QQQLQ 3681

RESULT 14 S25365

CYC8 protein - yeast (*Saccharomyces cerevisiae*)
 N;Alternate names: glucose repression mediator; protein YBR0908; protein YBR112c; SSN6 protein
 C;Species: *Saccharomyces cerevisiae*
 C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jan-2000
 C;Accession: S25365; S48277; S45980; S25404; S25405; A30906; S44692
 R;Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann, H. Yeast 8, 397-408, 1992
 A;Title: Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excision repair gene RAD16 located in this region belongs to a novel group of double-finger proteins.
 A;Reference number: S25364; MUID:92327848; PMID:1626431
 A;Accession: S25365
 A;Molecule type: DNA
 A;Residues: 1-966 <MAN>
 A;Cross-references: EMBL:X66247; NID:g3548; PIDN:CAA46973.1; PID:g3550
 R;Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann, H. Yeast 10, 1363-1381, 1994
 A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
 A;Reference number: S48255; MUID:95208357; PMID:7900426
 A;Accession: S48277
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-966 <MAW>

A;Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55615.1; PID:g476068
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994
R;Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45980
A;Molecule type: DNA
A;Residues: 1-966 <FE2>
A;Cross-references: EMBL:Z35981; NID:g536449; PIDN:CAA85069.1; PID:g536450;
MIPS:YBR112c
R;Schultz, J.; Carlson, M.
Mol. Cell. Biol. 7, 3637-3645, 1987
A;Title: Molecular analysis of SSN6, a gene functionally related to the SNF1
protein kinase of *Saccharomyces cerevisiae*.
A;Reference number: S25404; MUID:88065502; PMID:3316983
A;Accession: S25404
A;Molecule type: DNA
A;Residues: 1-546,'K',548-966 <SCH>
A;Cross-references: EMBL:M17826; NID:g172725; PIDN:AAA35103.1; PID:g172726
R;Trumbly, R.J.
Gene 73, 97-111, 1988
A;Title: Cloning and characterization of the CYC8 gene mediating glucose
repression in yeast.
A;Reference number: S25405; MUID:89211964; PMID:2854095
A;Accession: S25405
A;Molecule type: DNA
A;Residues: 1-546,'K',548-966 <TRU>
A;Cross-references: EMBL:M23440; NID:g171349; PIDN:AAA34545.1; PID:g171350
C;Genetics:
A;Gene: SGD:CYC8; SSN6; CRT8
A;Cross-references: SGD:S0000316; MIPS:YBR112c
A;Map position: 2R
C;Function:
A;Description: required for complete derepression of ICL1; required for
repression of SUC2 at high glucose levels and for induction of SUC2 at low
glucose levels
C;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide
repeat homology
C;Keywords: nucleus; transcription regulation
F;224-257/Domain: tetratricopeptide repeat homology <TT1>
F;262-295/Domain: tetratricopeptide repeat homology <TT2>
F;296-329/Domain: tetratricopeptide repeat homology <TT3>
F;330-363/Domain: tetratricopeptide repeat homology <TT4>
F;365-398/Domain: tetratricopeptide repeat homology <TT5>

Query Match 51.3%; Score 191; DB 2; Length 966;
Best Local Similarity 88.6%; Pred. No. 5.9e-09;
Matches 39; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      25  QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLQP 68
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      547  QAQAQAQAQAQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLQP 590
```

RESULT 15
S54522

hypothetical protein YMR164c - yeast (*Saccharomyces cerevisiae*)
 N;Alternate names: hypothetical protein YM8520.13c
 C;Species: *Saccharomyces cerevisiae*
 C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
 C;Accession: S54522; S54609
 R;Hunt, S.; Bowman, S.
 submitted to the EMBL Data Library, May 1995
 A;Reference number: S54510
 A;Accession: S54522
 A;Molecule type: DNA
 A;Residues: 1-758 <HUN>
 A;Cross-references: GB:Z49705; EMBL:Z49700; NID:g825556; PIDN:CAA89800.1;
 PID:g825569; EMBL:Z49705; MIPS:YMR164c
 A;Experimental source: strain AB972
 C;Genetics:
 A;Gene: SGD:MSS11
 A;Cross-references: SGD:S0004774; MIPS:YMR164c
 A;Map position: 13R

Query Match 51.1%; Score 190; DB 2; Length 758;
 Best Local Similarity 60.5%; Pred. No. 5.8e-09;
 Matches 46; Conservative 2; Mismatches 12; Indels 16; Gaps 2;

```

Qy      8 ATLEKLMKAFESLKSFQQQQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQ 52
          ||: | | | |:| | ||          | ||||| | | | | | | | | | | | | | |
Db      257 ATI-NLHKHFNDLQSPAQPQQSSQQQIQQPQHQPQHQPQQQQQQQQQQQQQQQQQQ 315

Qy      53 QQQQQQQQQQQQQQLQP 68
          ||||| | | | | |
Db      316 QQQQQQQQHQQQQQTP 331
  
```

Search completed: March 12, 2004, 15:41:46
 Job time : 12.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:39:10 ; Search time 30.1765 Seconds
(without alignments)
531.793 Million cell updates/sec

Title: US-09-620-955B-11
Perfect score: 372
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 76

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score Match Length DB ID	Description

1	293	78.8	171	14	US-10-077-584-4	Sequence 4, Appli
2	211.5	56.9	338	9	US-09-933-638A-12	Sequence 12, Appl
3	211.5	56.9	339	15	US-10-116-275-184	Sequence 184, App
4	211.5	56.9	371	9	US-09-849-243-16	Sequence 16, Appl
5	207	55.6	910	10	US-09-086-436-31	Sequence 31, Appl
6	205	55.1	63	14	US-10-077-584-6	Sequence 6, Appli
7	204	54.8	91	15	US-10-354-246-1	Sequence 1, Appli
8	197.5	53.1	4952	15	US-10-051-874-56	Sequence 56, Appl
9	197.5	53.1	5008	15	US-10-051-874-166	Sequence 166, App
10	197.5	53.1	5159	15	US-10-085-198-112	Sequence 112, App
11	197.5	53.1	5262	15	US-10-051-874-165	Sequence 165, App
12	197.5	53.1	5262	15	US-10-051-874-167	Sequence 167, App
13	197	53.0	97	9	US-09-864-761-35499	Sequence 35499, A
14	191	51.3	966	9	US-09-801-368-372	Sequence 372, App
15	190	51.1	87	14	US-10-215-432-27	Sequence 27, Appl
16	190	51.1	758	9	US-09-801-368-224	Sequence 224, App
17	190	51.1	1543	9	US-09-904-987-7	Sequence 7, Appli
18	182	48.9	80	14	US-10-177-725-14	Sequence 14, Appl
19	182	48.9	429	9	US-09-987-107-34	Sequence 34, Appl
20	181	48.7	2150	13	US-10-135-322-17	Sequence 17, Appl
21	179	48.1	4823	15	US-10-051-874-169	Sequence 169, App
22	173.5	46.6	398	15	US-10-374-780A-2358	Sequence 2358, Ap
23	173	46.5	314	14	US-10-317-832-13	Sequence 13, Appl
24	172	46.2	467	9	US-09-416-384A-7	Sequence 7, Appli
25	172	46.2	756	14	US-10-100-252-20	Sequence 20, Appl
26	172	46.2	756	15	US-10-369-493-1967	Sequence 1967, Ap
27	172	46.2	1138	14	US-10-074-475-194	Sequence 194, App
28	171.5	46.1	1070	9	US-09-735-367B-6	Sequence 6, Appli
29	171.5	46.1	2005	9	US-09-735-367B-3	Sequence 3, Appli
30	171.5	46.1	2063	9	US-09-735-367B-2	Sequence 2, Appli
31	171	46.0	228	13	US-10-029-180-16	Sequence 16, Appl
32	171	46.0	1572	15	US-10-116-275-179	Sequence 179, App
33	170	45.7	1420	14	US-10-379-616-4	Sequence 4, Appli
34	168	45.2	1403	13	US-10-108-605-93	Sequence 93, Appl
35	167.5	45.0	351	15	US-10-264-049-2693	Sequence 2693, Ap
36	166	44.6	623	15	US-10-464-939-12	Sequence 12, Appl
37	165	44.4	702	14	US-10-161-051-18	Sequence 18, Appl
38	163	43.8	1955	14	US-10-293-504-3	Sequence 3, Appli
39	162.5	43.7	401	9	US-09-987-107-36	Sequence 36, Appl
40	161.5	43.4	423	15	US-10-302-267-30	Sequence 30, Appl
41	161.5	43.4	423	15	US-10-374-780A-2388	Sequence 2388, Ap
42	160.5	43.1	286	9	US-09-801-368-188	Sequence 188, App
43	160	43.0	406	15	US-10-369-493-3147	Sequence 3147, Ap
44	159	42.7	79	9	US-09-864-761-33972	Sequence 33972, A
45	159	42.7	326	14	US-10-029-386-32987	Sequence 32987, A

ALIGNMENTS

RESULT 1

US-10-077-584-4

; Sequence 4, Application US/10077584

; Publication No. US20030073610A1

; GENERAL INFORMATION:

; APPLICANT: LINDQUIST, SUSAN


```
; APPLICANT: KROBITSCH, SYLVIA
; APPLICANT: OUTEIRO, TIAGO F.
; TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
; FILE REFERENCE: ARCD:367US
; CURRENT APPLICATION NUMBER: US/10/077,584
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269,157
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-584-4
```

```
Query Match          78.8%; Score 293; DB 14; Length 171;
Best Local Similarity 98.4%; Pred. No. 1.9e-22;
Matches 60; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60

Qy      67 Q 67
          |
Db      61 Q 61
```

RESULT 2

US-09-933-638A-12

```
; Sequence 12, Application US/09933638A
; Patent No. US20020160952A1
; GENERAL INFORMATION:
; APPLICANT: Kazantsev, Aleksey G.
; APPLICANT: Thompson, Leslie M.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: INHIBITION OF PROTEIN-PROTEIN INTERACTION
; FILE REFERENCE: 01997-289001
; CURRENT APPLICATION NUMBER: US/09/933,638A
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: US 60/226,502
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-933-638A-12
```

```
Query Match          56.9%; Score 211.5; DB 9; Length 338;
Best Local Similarity 67.7%; Pred. No. 6.9e-14;
Matches 44; Conservative 8; Mismatches 12; Indels 1; Gaps 1;
```

```
Qy      1 LVPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60
          ::| |: |: : || ::||:||||||||||||||||||||||||||
```

```

Db          31 MMPYGTGLTPQPIQNT-NSLSILEEQRRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 89
Qy          61 QQQQQ 65
              |||||
Db          90 QOOOO 94

```

RESULT 3

```

US-10-116-1275-184
; Sequence 184, Application US/10116275
; Publication No. US20030211476A1
; GENERAL INFORMATION:
; APPLICANT: Elan Pharmaceutical Technology
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Brayden, David
; APPLICANT: Byrne, Daragh
; APPLICANT: Lambkin, Imelda
; APPLICANT: Higgins, Lisa
; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and
Methods and
; TITLE OF INVENTION: Compositions Targeting Peyer's Patches and M Cell
Receptors
; FILE REFERENCE: E1067/20087
; CURRENT APPLICATION NUMBER: US/10/116,275
; CURRENT FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 349
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 184
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-275-184

```

Query Match 56.9%; Score 211.5; DB 15; Length 339;
Best Local Similarity 67.7%; Pred. No. 6.9e-14;
Matches 44; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

```

Qy      1 LVPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60
      ::| |: | : :   ||   ::||:|||||||||||||||||||||||||||||||||
Db      31 MMPYGTGLTPQPIQNT-NSLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 89

Qy      61 QQQQQ 65
      |||||
Db      90 QQQQQ 94

```

RESULT 4

US-09-849-243-16
; Sequence 16, Application US/09849243
; Patent No. US20020157127A1
; GENERAL INFORMATION:
; APPLICANT: Kirschbaum, Bernd
; Berglund, Erick
; Meisterernst, Michael
; Polites, Greg
; TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
; COMPLEXES FROM TRANSGENIC

```

;                               NON-HUMAN ANIMALS
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: HELLER, EHRMAN, WHITE & McAULIFFE
;       STREET: 1666 K Street, N.W., Suite 300
;       CITY: Washington
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 20006
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/849,243
;       FILING DATE: 07-May-2001
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Granados, Patricia D.
;       REGISTRATION NUMBER: 33,683
;       REFERENCE/DOCKET NUMBER: 38005-0148
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (202)912-2000
;       TELEFAX: (202)912-2020
;   INFORMATION FOR SEQ ID NO: 16:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 371 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-849-243-16

```

```

Query Match          56.9%;  Score 211.5;  DB 9;  Length 371;
Best Local Similarity 67.7%;  Pred. No. 7.6e-14;
Matches   44;  Conservative   8;  Mismatches  12;  Indels    1;  Gaps    1;

```

```

Qy      1 LVPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60
      ::| |: | : :   ||  ::||:|||||:|||||:|||||:|||||:|||||:
Db      63 MPPYGTGLTPQPIQNT-NSLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 121

Qy      61 QQQQQ 65
      |||||
Db      122 QQQQQ 126

```

RESULT 5

US-09-086-436-31

```

; Sequence 31, Application US/09086436
; Publication No. US20030118988A1
; GENERAL INFORMATION:
;   APPLICANT: Kandel, Eric R.
;   APPLICANT: Santoro, Bina
;   APPLICANT: Bartsch, Dusan
;   APPLICANT: Siegelbaum, Steven
;   APPLICANT: Tibbs, Gareth
;   APPLICANT: Grant, Seth

```



```
; GENERAL INFORMATION:
; APPLICANT: Khoshnan, Ali
; APPLICANT: Patterson, Paul H.
; TITLE OF INVENTION: ANTIBODIES THAT BIND TO AN EPITOPE OF
; TITLE OF INVENTION: THE HUNTINGTON'S DISEASE PROTEIN
; FILE REFERENCE: CALTE.012A
; CURRENT APPLICATION NUMBER: US/10/354,246
; CURRENT FILING DATE: 2003-01-28
; PRIOR APPLICATION NUMBER: 60/353,032
; PRIOR FILING DATE: 2001-01-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-354-246-1
```

```
Query Match          54.8%; Score 204; DB 15; Length 91;
Best Local Similarity 72.6%; Pred. No. 1e-13;
Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
```

```
Qy      7 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQL 66
          |||||||||||||||||||||||||||||||||||
Db      1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPPQA 60

Qy      67 QP 68
          ||
Db      61 QP 62
```

RESULT 8

US-10-051-874-56

```
; Sequence 56, Application US/10051874
; Publication No. US20040005557A1
```

; GENERAL INFORMATION:

```
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
```

```

; APPLICANT:  Smithson, Glennnda
; APPLICANT:  Baumgartner, Jason C
; APPLICANT:  Herrman, John L
; APPLICANT:  Peyman, John A
; APPLICANT:  Gorman, Linda
; APPLICANT:  Mezes, Peter D
; APPLICANT:  Kekuda, Ramesh
; APPLICANT:  Taupier Jr, Raymond J
; APPLICANT:  Gerlach, Valerie
; APPLICANT:  Grosse, William M
; APPLICANT:  Liu, Xiaohong
; APPLICANT:  Ellerman, Karen
; APPLICANT:  Rothenberg, Mark
; APPLICANT:  Stone, David J
; APPLICANT:  Burgess, Catherine E
; TITLE OF INVENTION:  PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION:  USING THE SAME
; FILE REFERENCE:  21402-245
; CURRENT APPLICATION NUMBER:  US/10/051,874
; CURRENT FILING DATE:  2002-09-25
; PRIOR APPLICATION NUMBER:  60/268,595
; PRIOR FILING DATE:  2001-02-14
; PRIOR APPLICATION NUMBER:  60/325,306
; PRIOR FILING DATE:  2001-09-27
; PRIOR APPLICATION NUMBER:  60/262,587
; PRIOR FILING DATE:  2001-01-18
; PRIOR APPLICATION NUMBER:  60/272,409
; PRIOR FILING DATE:  2001-02-28
; PRIOR APPLICATION NUMBER:  60/262,454
; PRIOR FILING DATE:  2001-01-18
; PRIOR APPLICATION NUMBER:  60/276,777
; PRIOR FILING DATE:  2001-03-16
; PRIOR APPLICATION NUMBER:  60/291,672
; PRIOR FILING DATE:  2001-05-17
; PRIOR APPLICATION NUMBER:  60/330,336
; PRIOR FILING DATE:  2001-10-18
; PRIOR APPLICATION NUMBER:  60/265,530
; PRIOR FILING DATE:  2001-01-31
; PRIOR APPLICATION NUMBER:  60/261,376
; PRIOR FILING DATE:  2001-01-16
; NUMBER OF SEQ ID NOS:  269
; SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 56
;   LENGTH:  4952
;   TYPE:  PRT
;   ORGANISM:  Homo sapiens
US-10-051-874-56

```

```

Query Match          53.1%;  Score 197.5;  DB 15;  Length 4952;
Best Local Similarity 66.2%;  Pred. No. 2.8e-11;
Matches  43;  Conservative  6;  Mismatches  15;  Indels  1;  Gaps  1;

```

```

QY          3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 62
             | ||:  |::  :  :  :  | |||||  ||||  |||  |||||  |||||  |||
Db          3313 PMGSLQQQLQQ-QQQQLQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQQQLQQQQQQQLQQQ 3371

QY          63 QQQLQ 67

```

Db |||||
 3372 QQQLQ 3376

RESULT 9

US-10-051-874-166

; Sequence 166, Application US/10051874

; Publication No. US20040005557A1

; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Alsobrook II, John P

; APPLICANT: Colman, Steven D

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Boldog, Ferenc

; APPLICANT: Vernet, Corine AM

; APPLICANT: Li, Li

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Casman, Stacie J

; APPLICANT: Guo, Xiaojia Sasha

; APPLICANT: Edinger, Shlomit R

; APPLICANT: MacDougall, John R

; APPLICANT: Malyankar, Uriel M

; APPLICANT: Patturajan, Meera

; APPLICANT: Shimkets, Richard A

; APPLICANT: Pena, Carol EA

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Zerhusen, Bryan D

; APPLICANT: Millet, Isabelle

; APPLICANT: Miller, Charles E

; APPLICANT: Lepley, Denise M

; APPLICANT: Smithson, Glennnda

; APPLICANT: Baumgartner, Jason C

; APPLICANT: Herrman, John L

; APPLICANT: Peyman, John A

; APPLICANT: Gorman, Linda

; APPLICANT: Mezes, Peter D

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Gerlach, Valerie

; APPLICANT: Grosse, William M

; APPLICANT: Liu, Xiaohong

; APPLICANT: Ellerman, Karen

; APPLICANT: Rothenberg, Mark

; APPLICANT: Stone, David J

; APPLICANT: Burgess, Catherine E

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-245

; CURRENT APPLICATION NUMBER: US/10/051,874

; CURRENT FILING DATE: 2002-09-25

; PRIOR APPLICATION NUMBER: 60/268,595

; PRIOR FILING DATE: 2001-02-14

; PRIOR APPLICATION NUMBER: 60/325,306

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/262,587

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,409

```

; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 166
; LENGTH: 5008
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-051-874-166

```

```

Query Match          53.1%; Score 197.5; DB 15; Length 5008;
Best Local Similarity 66.2%; Pred. No. 2.8e-11;
Matches 43; Conservative 6; Mismatches 15; Indels 1; Gaps 1;

```

```

Qy          3 PRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 62
| ||: |:: : : : ||||| |||| | || | ||||| ||||| |||
Db          3364 PMGSLQQQLQQ-QQQQLQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQQQLQQQ 3422

Qy          63 QQQLQ 67
| ||||
Db          3423 QQQLQ 3427

```

RESULT 10

```

US-10-085-198-112
; Sequence 112, Application US/10085198
; Publication No. US20040009907A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-279
; CURRENT APPLICATION NUMBER: US/10/085,198
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/271,646
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/276,401
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,981
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,858
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/271,840
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/277,324
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/286,096

```



```

; PRIOR FILING DATE: 2001-04-21
; PRIOR APPLICATION NUMBER: 60/299,695
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/315,614
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/272,405
; PRIOR FILING DATE: 2001-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 112
; LENGTH: 5159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-085-198-112

```

```

Query Match          53.1%; Score 197.5; DB 15; Length 5159;
Best Local Similarity 66.2%; Pred. No. 2.9e-11;
Matches 43; Conservative 6; Mismatches 15; Indels 1; Gaps 1;

```

```

Qy          3 PRGSMATLEKLMKAFESLKSFOQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 62
| ||: |:: : : : | ||||| ||||| |||| | ||||| | ||||| |||
Db          3313 PMGSLQQLQQ-QQQLQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQQQLQQQQQLQQ 3371

Qy          63 QQQLQ 67
| ||||
Db          3372 QQQLQ 3376

```

RESULT 11

```

US-10-051-874-165
; Sequence 165, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennda
; APPLICANT: Baumgartner, Jason C

```

```

; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 165
; LENGTH: 5262
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-051-874-165

```

```

Query Match          53.1%; Score 197.5; DB 15; Length 5262;
Best Local Similarity 66.2%; Pred. No. 2.9e-11;
Matches 43; Conservative 6; Mismatches 15; Indels 1; Gaps 1;

```

```

Qy      3 PRGSMATLEKLMKAFESLKSFOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO 62
        | ||: |:: : : : | ||||| |||| | || | || || || || || || || || ||
Db      3618 PMGSLQQ-LQQ-QQQ-LQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQLQQQQQLQQ 3676

Qy      63 QQQLQ 67
        |||||
Db      3677 QQQLQ 3681

```

RESULT 12

US-10-051-874-167

; Sequence 167, Application US/10051874

; Publication No. US20040005557A1

; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Alsobrook II, John P

; APPLICANT: Colman, Steven D

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Boldog, Ferenc

; APPLICANT: Vernet, Corine AM

; APPLICANT: Li, Li

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Casman, Stacie J

; APPLICANT: Guo, Xiaojia Sasha

; APPLICANT: Edinger, Shlomit R

; APPLICANT: MacDougall, John R

; APPLICANT: Malyankar, Uriel M

; APPLICANT: Patturajan, Meera

; APPLICANT: Shimkets, Richard A

; APPLICANT: Pena, Carol EA

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Zerhusen, Bryan D

; APPLICANT: Millet, Isabelle

; APPLICANT: Miller, Charles E

; APPLICANT: Lepley, Denise M

; APPLICANT: Smithson, Glennnda

; APPLICANT: Baumgartner, Jason C

; APPLICANT: Herrman, John L

; APPLICANT: Peyman, John A

; APPLICANT: Gorman, Linda

; APPLICANT: Mezes, Peter D

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Gerlach, Valerie

; APPLICANT: Grosse, William M

; APPLICANT: Liu, Xiaohong

; APPLICANT: Ellerman, Karen

; APPLICANT: Rothenberg, Mark

; APPLICANT: Stone, David J

; APPLICANT: Burgess, Catherine E

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-245

; CURRENT APPLICATION NUMBER: US/10/051,874

; CURRENT FILING DATE: 2002-09-25

; PRIOR APPLICATION NUMBER: 60/268,595

; PRIOR FILING DATE: 2001-02-14

; PRIOR APPLICATION NUMBER: 60/325,306

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/262,587

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,409

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18
 ; PRIOR APPLICATION NUMBER: 60/276,777
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/291,672
 ; PRIOR FILING DATE: 2001-05-17
 ; PRIOR APPLICATION NUMBER: 60/330,336
 ; PRIOR FILING DATE: 2001-10-18
 ; PRIOR APPLICATION NUMBER: 60/265,530
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/261,376
 ; PRIOR FILING DATE: 2001-01-16
 ; NUMBER OF SEQ ID NOS: 269
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 167
 ; LENGTH: 5262
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-051-874-167

Query Match 53.1%; Score 197.5; DB 15; Length 5262;
 Best Local Similarity 66.2%; Pred. No. 2.9e-11;
 Matches 43; Conservative 6; Mismatches 15; Indels 1; Gaps 1;

Qy 3 PRGSMATLEKLMKAFESLKSFQQ 62
 | |:: |:: : : : | | | | | | | | | | | | | | | | | | | | | | | |
 Db 3618 PMGSLQQLQQ-QQQLQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQQQLQQQQQQQLQQQ 3676
 Qy 63 QQQLQ 67
 | | | |
 Db 3677 QQQLQ 3681

RESULT 13

US-09-864-761-35499

; Sequence 35499, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
 USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

```

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35499
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009954.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 47
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 53
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 69
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 27
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 21
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 45
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 29
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 33
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 21
; OTHER INFORMATION: EST_HUMAN HIT: BE260046.1, EVALUE 3.00e-14
; OTHER INFORMATION: SWISSPROT HIT: P53360, EVALUE 3.00e-15
US-09-864-761-35499

```

```

Query Match          53.0%; Score 197; DB 9; Length 97;
Best Local Similarity 53.7%; Pred. No. 5.7e-13;
Matches 44; Conservative 11; Mismatches 19; Indels 8; Gaps 2;

```

```

Qy      1 LVPRGSMATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60
        ::| |: | : :   ||  ::||:|||||:|||||:|||||:|||||:|||||:
Db      13 MMPYGTGLTPQPIQNT-NSLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 71

Qy      61 QQ-----QQQLQPGSTRAAA 75

```

Db 72 QQAVAAA^{||}AVOOST^{||}S^{:|:}OOATOGTS[:] 93

RESULT 14

```

US-09-801-368-372
; Sequence 372, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in
Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 372
; LENGTH: 966
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-372

```

Query Match 51.3%; Score 191; DB 9; Length 966;
Best Local Similarity 88.6%; Pred. No. 2.4e-11;
Matches 39; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 25 QQLQP 68
 | | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 QAQAQAQAQAQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLQP 590

RESULT 15

US-10-215-432-27
; Sequence 27, Application US/10215432
; Publication No. US20030109476A1
; GENERAL INFORMATION:
; APPLICANT: Eric B. Kmiec
; APPLICANT: Hetal Parekh-Olmedo
; TITLE OF INVENTION: Composition and methods for the
; TITLE OF INVENTION: prevention and treatment of Huntington's disease

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:34:19 ; Search time 37.2549 Seconds
(without alignments)
643.657 Million cell updates/sec

Title: US-09-620-955B-11
Perfect score: 372
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 76

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriaphage:*
17: sp_archaeophages:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description

1	229	61.6	1156	5	Q86HG5	Q86hg5 dictyosteli
2	225	60.5	618	16	Q87G62	Q87g62 vibrio para
3	224.5	60.3	739	11	Q7TPU6	Q7tpu6 mus musculu
4	224	60.2	646	16	Q9KMZ5	Q9kmz5 vibrio chol
5	223	59.9	1080	5	Q86KL1	Q86kl1 dictyosteli
6	222.5	59.8	1163	5	Q869M3	Q869m3 dictyosteli
7	222	59.7	1693	5	Q86JI7	Q86ji7 dictyosteli
8	221	59.4	650	16	Q8D3X1	Q8d3x1 vibrio vuln
9	221	59.4	2472	5	Q8MXN1	Q8mxn1 dictyosteli
10	220	59.1	522	13	O42323	O42323 coturnix co
11	219.5	59.0	1297	5	Q8SSS5	Q8sss5 dictyosteli
12	219	58.9	827	5	Q86KD2	Q86kd2 dictyosteli
13	219	58.9	1502	5	Q8IS10	Q8is10 dictyosteli
14	218.5	58.7	1531	5	Q86GH1	Q86gh1 drosophila
15	218	58.6	149	4	Q8NFT3	Q8nft3 homo sapien
16	218	58.6	557	3	Q8X0K6	Q8x0k6 neurospora
17	218	58.6	1811	5	Q8IJD3	Q8ijd3 plasmodium
18	217	58.3	1204	5	Q8T134	Q8t134 dictyosteli
19	217	58.3	1832	3	Q8TGH8	Q8tgh8 podospora a
20	217	58.3	2123	5	Q9U9S7	Q9u9s7 dictyosteli
21	216.5	58.2	1329	5	Q86AA2	Q86aa2 dictyosteli
22	216	58.1	680	5	Q86AM9	Q86am9 dictyosteli
23	215.5	57.9	1969	5	O15763	O15763 dictyosteli
24	215	57.8	856	5	Q8T151	Q8t151 dictyosteli
25	215	57.8	1457	5	O44011	O44011 dictyosteli
26	214	57.5	652	5	Q8T2S4	Q8t2s4 dictyosteli
27	214	57.5	1212	5	Q86AF2	Q86af2 dictyosteli
28	213.5	57.4	602	5	Q86GH6	Q86gh6 drosophila
29	213.5	57.4	809	13	Q7ZVN7	Q7zvn7 brachydanio
30	213.5	57.4	1330	5	Q86GH2	Q86gh2 drosophila
31	213.5	57.4	1537	5	Q86GH5	Q86gh5 drosophila
32	213	57.3	218	6	Q8MHX3	Q8mhx3 pan troglod
33	213	57.3	646	5	Q8MKN4	Q8mnk4 dictyosteli
34	213	57.3	716	6	Q8MJA0	Q8mja0 pan troglod
35	213	57.3	716	6	Q8HZ00	Q8hz00 pan paniscu
36	213	57.3	3417	5	Q86J15	Q86j15 dictyosteli
37	212	57.0	3770	5	Q869R6	Q869r6 dictyosteli
38	211.5	56.9	151	4	Q7Z6S4	Q7z6s4 homo sapien
39	211.5	56.9	208	4	Q7Z6S5	Q7z6s5 homo sapien
40	211.5	56.9	1918	5	Q86AF5	Q86af5 dictyosteli
41	211	56.7	217	4	Q8N0W2	Q8n0w2 homo sapien
42	211	56.7	222	4	Q8NFQ4	Q8nfq4 homo sapien
43	211	56.7	365	4	Q8NFQ1	Q8nfq1 homo sapien
44	211	56.7	415	4	Q8NFQ3	Q8nfq3 homo sapien
45	211	56.7	431	4	Q8N6B6	Q8n6b6 homo sapien

ALIGNMENTS

RESULT 1

Q86HG5

ID Q86HG5 PRELIMINARY; PRT; 1156 AA.

AC Q86HG5;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Similar to similar to Uba2p; Ubalp.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=22092622; PubMed=12097910;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
 RL Nature 418:79-85(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Baumgart C.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC117072; AA052629.1; -.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0004839; F:ubiquitin activating enzyme activity; IEA.
 DR GO; GO:0006512; P:ubiquitin cycle; IEA.
 DR InterPro; IPR009036; MoeB.
 DR InterPro; IPR000205; NAD_BS.
 DR InterPro; IPR000594; ThiF_domain.
 DR InterPro; IPR000127; UBact_repeat.
 DR InterPro; IPR000011; Ugtin-activ_enz.
 DR Pfam; PF00899; ThiF; 2.
 DR Pfam; PF02134; UBACT; 2.
 DR TIGRFAMs; TIGR01408; Ube1; 1.
 SQ SEQUENCE 1156 AA; 134093 MW; E949F2A47DA46A86 CRC64;

Query Match 61.6%; Score 229; DB 5; Length 1156;
 Best Local Similarity 62.2%; Pred. No. 1e-15;
 Matches 51; Conservative 6; Mismatches 9; Indels 16; Gaps 2;

Qy 1 LVPRGSMATLEKLMKAFESL-----KSFQQQQQQQQQQQQQQQQQQQQQQ 46
 ::| ::|| :: | || | |||||
 Db 946 IIP--AIATTSVIAGFVSLELIKVLSSNYYQFKQSQQQQQQQQQQQQQQQQQQQQ 1003
 Qy 47 QQQQQQQQQQQQQQQQQQQQQQLQP 68
 |||||
 Db 1004 QQQQQQQQQQQQQQQQQQQQQQP 1025

RESULT 2

Q87G62

ID Q87G62 PRELIMINARY; PRT; 618 AA.
 AC Q87G62;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN VPA1455.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.

DR PIR; D82493; D82493.
 DR TIGR; VCA0171; -.
 DR InterPro; IPR001440; TPR.
 DR InterPro; IPR008941; TPR-like.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00515; TPR; 1.
 DR SMART; SM00327; VWA; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 646 AA; 71064 MW; 87E17761CFE38CE6 CRC64;

Query Match 60.2%; Score 224; DB 16; Length 646;
 Best Local Similarity 80.7%; Pred. No. 2e-15;
 Matches 46; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 20 LKSFQQQLPGSTRAAAS 76
 :|: ||||| : : ||
 Db 440 VKAAQQQDSSSGAS 496

RESULT 5

Q86KL1

ID Q86KL1 PRELIMINARY; PRT; 1080 AA.
 AC Q86KL1;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=22092622; PubMed=12097910;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tunggal B., Cox E., Quail M.A., Platzner M., Rosenthal A., Noegel A.A.;
 RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
 RL Nature 418:79-85(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Baumgart C.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC117070; AA051011.1; -.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00097; zf-C3HC4; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00518; ZF_RING_1; 1.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1080 AA; 126555 MW; 432C39D84C26ED29 CRC64;

Query Match 59.9%; Score 223; DB 5; Length 1080;
 Best Local Similarity 86.8%; Pred. No. 4.1e-15;
 Matches 46; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=22092622; PubMed=12097910;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tunggal B., Cox E., Quail M.A., Platzner M., Rosenthal A., Noegel A.A.;
 RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
 RL Nature 418:79-85(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Baumgart C.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC116982; AAO51575.1; -.
 SQ SEQUENCE 1693 AA; 194003 MW; 178D25E074974B10 CRC64;

Query Match 59.7%; Score 222; DB 5; Length 1693;
 Best Local Similarity 69.1%; Pred. No. 7.8e-15;
 Matches 47; Conservative 5; Mismatches 6; Indels 10; Gaps 1;

Qy 10 LEKLMKAFES-----LKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 59
 :||::| : | : |||||
 Db 1607 IEKMLKQQQQQQQLQQQYQQHLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 1666
 Qy 60 QQQQQQLQ 67
 ||||| |
 Db 1667 QQQQQQQQ 1674

RESULT 8

Q8D3X1

ID Q8D3X1 PRELIMINARY; PRT; 650 AA.
 AC Q8D3X1;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE TPR repeat containing protein.
 GN VV21562.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.E.;
 RT "Complete genome sequence of Vibrio vulnificus CMCP6."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AE016813; AAO08425.1; -.
 DR InterPro; IPR001440; TPR.
 DR InterPro; IPR008941; TPR-like.
 DR InterPro; IPR002035; VWF_A.

Db 259 QQQQQQQ 265

RESULT 10

O42323

ID O42323 PRELIMINARY; PRT; 522 AA.
AC O42323;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE QMEF2D protein.
GN QMEF2D.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20461869; PubMed=10899598;
RA Xue Z.G., Xue X.J., Roncier B., Chamagne A.M., Portier M.M.;
RT "Isolation of quail qMEF2D gene and its expression pattern in the
RT developing central nervous system."
RL Biochim. Biophys. Acta 1492:543-547(2000).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC FACTORS.
DR EMBL; AJ002238; CAA05282.1; -.
DR HSSP; P11831; 1SRS.
DR TRANSFAC; T03821; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR002100; TF_MADSbox.
DR Pfam; PF00319; SRF-TF; 1.
DR PRINTS; PR00404; MADSDOMAIN.
DR SMART; SM00432; MADS; 1.
DR PROSITE; PS00350; MADS_BOX_1; 1.
DR PROSITE; PS50066; MADS_BOX_2; 1.
KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.
SQ SEQUENCE 522 AA; 57615 MW; F51726DCDD95DC99 CRC64;

Query Match 59.1%; Score 220; DB 13; Length 522;
Best Local Similarity 84.6%; Pred. No. 4.4e-15;
Matches 44; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 19 SLKSFQQQLQPGS 70
:: ::|||||
Db 357 NISAWQQQHLVPVS 408

RESULT 11

Q8SSS5

ID Q8SSS5 PRELIMINARY; PRT; 1297 AA.
AC Q8SSS5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)

RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC116956; AAO51124.1; -.
DR InterPro; IPR000270; OPR_PBl.
DR Pfam; PF00564; PBl; 1.
DR SMART; SM00666; PBl; 1.
SQ SEQUENCE 827 AA; 91758 MW; ED4ED9FCE2BA291A CRC64;

Query Match 58.9%; Score 219; DB 5; Length 827;
Best Local Similarity 73.0%; Pred. No. 8.4e-15;
Matches 46; Conservative 5; Mismatches 6; Indels 6; Gaps 1;

Qy 11 EKLMKAFESL-----KSFQQ 64
::: | ||: : |||||
Db 763 DEITKEIESVFLKQQQQKLQQ 822
Qy 65 QLQ 67
| |
Db 823 QQQ 825

RESULT 13

Q8IS10

ID Q8IS10 PRELIMINARY; PRT; 1502 AA.
AC Q8IS10;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleotide exchange factor RasGEF P.
GN GEFP.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Wilkins A., Szafranski K., Gloeckner G., Harrisingh M.,
RA Deenadayalan B., Mueller R., Eichinger L., Noegel A.A., Insall R.;
RT "The family of rasGEF genes in Dictyostelium discoideum.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY160105; AAN46885.1; -.
DR GO; GO:0005085; F:guanyl-nucleotide exchange factor activity; IEA.
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGRF_CDC25.
DR InterPro; IPR008937; Ras_GEF.

Query Match 58.9%; Score 219; DB 5; Length 1502;
Best Local Similarity 88.2%; Pred. No. 1.4e-14;
Matches 45; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

ID      Q86GH1          PRELIMINARY;          PRT;  1531 AA.
AC      Q86GH1;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Pol protein.
GN      POL.
OS      Drosophila virilis (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7244;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22531870; PubMed=12626755;
RA      Casacuberta E., Pardue M.L.;
RT      "Transposon telomeres are widely distributed in the Drosophila genus:
RT      TART elements in the virilis group.";
RL      Proc. Natl. Acad. Sci. U.S.A. 100:3363-3368(2003).
DR      EMBL; AY219709; AAO67564.1; -.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR      GO; GO:0006278; P:RNA dependent DNA replication; IEA.
DR      InterPro; IPR005135; Exo_endo_phos.
DR      InterPro; IPR000477; RVTse.
DR      Pfam; PF03372; Exo_endo_phos; 1.
DR      Pfam; PF00078; rvt; 1.
SQ      SEQUENCE      1531 AA;  177648 MW;  836575372A353376 CRC64;

```

Query Match 58.7%; Score 218.5; DB 5; Length 1531;
Best Local Similarity 60.8%; Pred. No. 1.6e-14;
Matches 48; Conservative 6; Mismatches 10; Indels 15; Gaps 1;

Qy 4 RGSMTLEKLMKAFF-----SLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQ 48
 | ||| : | :: : |||||

```

Db          1086 REGRITLLEELKLAIREQPLVLIQQLVLPRELAIQIYQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 1145
Qy          49  QQQQQQQQQQQQQQQQQQQQLQ 67
              |||||:|||||
Db          1146 QQQQQQQQQQQQEQQQQQQQ 1164

```

Q8NFT3

```

ID      Q8NFT3          PRELIMINARY;          PRT;      149 AA.
AC      Q8NFT3;
DT      01-OCT-2002 (TrEMBLrel. 22, Created)
DT      01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Forkhead transcription factor (Fragment).
GN      FOXP2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Frontal cortex;
RA      Bruce H.A., Margolis R.L.;
RT      "FOXP2: novel exons, splice variants, and CAG repeat length
RT      stability.";
RL      Hum. Genet. 0:0-0(2002).
DR      EMBL; AF454830; AAM60761.1; -.
FT      NON_TER      149      149
SQ      SEQUENCE      149 AA: 18064 MW: 9DB79972CC4F35FA CRC64:

```

Query Match 58.6%; Score 218; DB 4; Length 149;
Best Local Similarity 68.9%; Pred. No. 2.3e-15;
Matches 51; Conservative 1; Mismatches 10; Indels 12; Gaps 2;

```

Qy      8 ATLEKLMKAF-----ESLKSFQQQ-----QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 55
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     52 AALEKNSKWFIIQQQLQEFYKKQQEQHLHQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 111

Qy     56 QQQQQQQQQQQQLQPG 69
      | | | | | | | | | | | |
Db    112 QQQQQQQQQQQQOHPG 125

```

Search completed: March 12, 2004, 15:41:01
Job time : 38.2549 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2004, 15:22:04 ; Search time 8.94118 Seconds
(without alignments)
442.596 Million cell updates/sec

Title: US-09-620-955B-11
Perfect score: 372
Sequence: 1 LVPRGSMATLEKLMKAFESL.....QQQQQQQQQLQPGSTRAAAS 76

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query	Match	Length	ID	Description	
1	220	59.1	1177	1	SP97_DICDI	Q95zg3 dictyosteli	
2	213	57.3	716	1	FXP2_PANTR	Q8mja0 pan troglod	
3	213	57.3	905	1	SNF5_YEAST	P18480 saccharomyc	
4	211.5	56.9	339	1	TBP_HUMAN	P20226 homo sapien	
5	211	56.7	714	1	FXP2_MOUSE	P58463 mus musculu	
6	211	56.7	715	1	FXP2_HUMAN	O15409 homo sapien	
7	208	55.9	1905	1	TAGB_DICDI	P54683 dictyosteli	
8	207	55.6	910	1	HCN1_MOUSE	O88704 mus musculu	
9	204	54.8	3144	1	HD_HUMAN	P42858 homo sapien	
10	199.5	53.6	2212	1	T230_HUMAN	Q93074 homo sapien	
11	198.5	53.4	1023	1	CLOC_DROME	O61735 drosophila	
12	197.5	53.1	5262	1	MLL2_HUMAN	O14686 homo sapien	
13	197	53.0	705	1	FXP1_MOUSE	P58462 mus musculu	
14	195.5	52.6	3726	1	ABF1_MOUSE	Q61329 mus musculu	
15	194	52.2	1080	1	HDC_DROME	Q9n2m8 drosophila	
16	190	51.1	758	1	YM38_YEAST	Q03825 saccharomyc	
17	188.5	50.7	931	1	LUG_ARATH	Q9fuy2 arabidopsis	

18	187	50.3	966	1	SSN6_YEAST	P14922	saccharomyc
19	184	49.5	648	1	KAPC_DICDI	P34099	dictyosteli
20	184	49.5	2067	1	NCO6_MOUSE	Q9j119	m nuclear r
21	183	49.2	910	1	HCN1_RAT	Q9jkb0	rattus norv
22	182	48.9	429	1	APA4_MACFA	P33621	macaca fasc
23	182	48.9	1167	1	WC1_NEUCR	Q01371	neurospora
24	181	48.7	1516	1	NCO2_XENLA	Q9w705	xenopus lae
25	178.5	48.0	1596	1	MAM_DROME	P21519	drosophila
26	174.5	46.9	519	1	ELAV_DROVI	P23241	drosophila
27	173.5	46.6	398	1	PF21_ARATH	Q04088	arabidopsis
28	172	46.2	756	1	CBK1_YEAST	P53894	saccharomyc
29	172	46.2	1090	1	NIT4_NEUCR	P28349	neurospora
30	171.5	46.1	1556	1	PROS_DROVI	Q9u6a1	drosophila
31	171.5	46.1	2063	1	NCO6_HUMAN	Q14686	h nuclear r
32	170	45.7	1081	1	GALY_YEAST	P19659	saccharomyc
33	170	45.7	1424	1	NCO3_HUMAN	Q9y6q9	h nuclear r
34	169	45.4	313	1	THAB_HUMAN	Q96ek4	homo sapien
35	169	45.4	2038	1	FSH_DROME	P13709	drosophila
36	168	45.2	1319	1	MN1_HUMAN	Q10571	homo sapien
37	168	45.2	1403	1	PROS_DROME	P29617	drosophila
38	167.5	45.0	788	1	PCAP_HUMAN	Q96rn5	homo sapien
39	166	44.6	623	1	DSH_DROME	P51140	drosophila
40	164	44.1	525	1	NAB2_YEAST	P32505	saccharomyc
41	163.5	44.0	319	1	GDA5_WHEAT	P04725	triticum ae
42	163	43.8	313	1	GDA7_WHEAT	P04727	triticum ae
43	163	43.8	792	1	PCAP_MOUSE	Q924h2	mus musculu
44	162.5	43.7	401	1	APA4_PAPAN	Q28758	papio anubi
45	162.5	43.7	467	1	INVO_MOUSE	P48997	mus musculu

ALIGNMENTS

RESULT 1

SP97_DICDI

ID SP97_DICDI STANDARD; PRT; 1177 AA.

AC Q95ZG3;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Spindle pole body component 97 (Spc97) (DdSpc97).

GN SPC97.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

OX NCBI_TaxID=44689;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AX2;

RX MEDLINE=22012446; PubMed=12018385;

RA Daunderer C., Graf R.O.;

RT "Molecular analysis of the cytosolic Dictyostelium gamma-tubulin complex.";

RL Eur. J. Cell Biol. 81:175-184(2002).

CC -!- FUNCTION: May be involved in microtubule nucleation.

CC -!- SUBCELLULAR LOCATION: Centrosome, and also found in the cytoplasm.

CC -!- SIMILARITY: Belongs to the GCP family.

CC

RL Nature 418:869-872(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22412141; PubMed=12524352;
 RA Zhang J., Webb D.M., Podlaha O.;
 RT "Accelerated protein evolution and origins of human-specific features:
 RT Foxp2 as an example.";
 RL Genetics 162:1825-1835(2002).
 CC -!- FUNCTION: Transcriptional repressor that plays an important role
 CC in the specification and differentiation of lung epithelium. May
 CC play important roles in developing neural, gastrointestinal and
 CC cardiovascular tissues (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF512947; AAN03385.1; -.
 DR EMBL; AF515051; AAN03409.1; -.
 DR EMBL; AF515052; AAN03410.1; -.
 DR EMBL; AY143178; AAN60056.1; -.
 DR InterPro; IPR001766; TF_Fork_head.
 DR InterPro; IPR009058; Wing_hlx_DNA_bnd.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
 DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
 DR PROSITE; PS50039; FORK_HEAD_3; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein.
 FT ZN_FING 347 372 C2H2-TYPE.
 FT DNA_BIND 505 595 FORK-HEAD.
 FT DOMAIN 53 56 POLY-GLN.
 FT DOMAIN 123 126 POLY-GLN.
 FT DOMAIN 131 136 POLY-GLN.
 FT DOMAIN 152 191 POLY-GLN.
 FT DOMAIN 201 210 POLY-GLN.
 FT DOMAIN 224 232 POLY-GLN.
 SQ SEQUENCE 716 AA; 80061 MW; 3169A2786B42F79F CRC64;

Query Match 57.3%; Score 213; DB 1; Length 716;
 Best Local Similarity 61.4%; Pred. No. 1.4e-10;
 Matches 51; Conservative 6; Mismatches 12; Indels 14; Gaps 3;

Qy 1 LVPRGSMATLEK---LMKAFESLKSF---QQ-----QQQQQQQQQQQQQQQQQQQQQQ 46

```

      | |:   | |:   :|   :|: |   ||           |||||
Db    113 LSPQQLQALLQQQQAVMLQQQQLEFYKKQEQLHLQLLQQQQQQQQQQQQQQQQQQQQ 172
      47 QQQQQQQQQQQQQQQQQQQQLPG 69
      |||||
Db    173 QQQQQQQQQQQQQQQQQQQQHHPG 195

```

RESULT 3

SNF5 YEAST

```

ID SNF5_YEAST STANDARD; PRT; 905 AA.
AC P18480;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription regulatory protein SNF5 (SWI/SNF complex component SNF5)
DE (Transcription factor TYE4).
GN SNF5 OR TYE4 OR SWI10 OR YBR289W OR YBR2036.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MCY;
RX MEDLINE=91042489; PubMed=2233708;
RA Laurent B.C., Treitel M.A., Carlson M.;
RT "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and
RT proline-rich transcriptional activator that affects expression of a
RT broad spectrum of genes.";
RL Mol. Cell. Biol. 10:5616-5625(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=94378722; PubMed=8091861;
RA Holmstroem K., Brandt T., Kallesoe T.;
RT "The sequence of a 32,420 bp segment located on the right arm of
RT chromosome II from Saccharomyces cerevisiae.";
RL Yeast 10:S47-S62(1994).
CC -!- FUNCTION: Involved in transcriptional activation. The SWI/SNF
CC complex is required for the induced expression of a large number
CC of genes. This complex alters chromatin structure to facilitate
CC binding of gene-specific dedicated transcription factors.
CC -!- SUBUNIT: Component of the SWI/SNF global transcription activator
CC complex.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the SNF5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M36482; AAA35062.1; -.

```


RX MEDLINE=90326195; PubMed=2374612;
 RA Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,
 RA Roeder R.G.;
 RT "Highly conserved core domain and unique N terminus with presumptive
 RT regulatory motifs in a human TATA factor (TFIID).";
 RL Nature 346:387-390(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Griffiths C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP INTERACTION WITH NCOA6.
 RX MEDLINE=20036574; PubMed=10567404;
 RA Lee S.-K., Anzick S.L., Choi J.-E., Bubendorf L., Guan X.-Y.,
 RA Jung Y.-K., Kallioniemi O.P., Kononen J., Trent J.M., Azorsa D.,
 RA Jhun B.-H., Cheong J.H., Lee Y.C., Meltzer P.S., Lee J.W.;
 RT "A nuclear factor ASC-2, as a cancer-amplified transcriptional
 RT coactivator essential for ligand-dependent transactivation by nuclear
 RT receptors in vivo.";
 RL J. Biol. Chem. 274:34283-34293(1999).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337 IN COMPLEX WITH DNA.
 RX MEDLINE=96209823; PubMed=8643494;
 RA Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,
 RA Burley S.K.;
 RT "Crystal structure of a human TATA box-binding protein/TATA element
 RT complex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4862-4867(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339 IN COMPLEX WITH DNA.
 RX MEDLINE=96346176; PubMed=8757291;
 RA Juo Z.S., Chiu T.K., Leiberman P.M., Baikalov I., Berk A.J.,
 RA Dickerson R.E.;
 RT "How proteins recognize the TATA box.";
 RL J. Mol. Biol. 261:239-254(1996).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS) OF 159-337 IN COMPLEX WITH
 RP GTF2B AND DNA.
 RX MEDLINE=20086817; PubMed=10619841;
 RA Tsai F.T.F., Sigler P.B.;
 RT "Structural basis of preinitiation complex assembly on human pol II
 RT promoters.";
 RL EMBO J. 19:25-36(2000).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.62 ANGSTROMS) OF 159-339 IN COMPLEX WITH DR1;
 RP DRAP1 AND DNA.
 RX MEDLINE=21354312; PubMed=11461703;
 RA Kamada K., Shu F., Chen H., Malik S., Stelzer G., Roeder R.G.,
 RA Meisterernst M., Burley S.K.;
 RT "Crystal structure of negative cofactor 2 recognizing the TBP-DNA
 RT transcription complex.";
 RL Cell 106:71-81(2001).
 RN [10]
 RP POLYMORPHISM OF POLY-GLN REGION.
 RX MEDLINE=99415745; PubMed=10484774;
 RA Koide R., Kobayashi S., Shimohata T., Ikeuchi T., Maruyama M.,
 RA Saito M., Yamada M., Takahashi H., Tsuji S.;

RT "A neurological disease caused by an expanded CAG trinucleotide repeat
 RT in the TATA-binding protein gene: a new polyglutamine disease?";
 RL Hum. Mol. Genet. 8:2047-2053(1999).
 RN [11]
 RP POLYMORPHISM OF POLY-GLN REGION.
 RX MEDLINE=21214723; PubMed=11313753;
 RA Zuhlke C., Hellenbroich Y., Dalski A., Kononowa N., Hagenah J.,
 RA Vieregge P., Riess O., Klein C., Schwinger E.;
 RT "Different types of repeat expansion in the TATA-binding protein gene
 RT are associated with a new form of inherited ataxia.";
 RL Eur. J. Hum. Genet. 9:160-164(2001).

RN [12]
 RP POLYMORPHISM OF POLY-GLN REGION.
 RX MEDLINE=21341926; PubMed=11448935;
 RA Nakamura K., Jeong S.-Y., Uchihara T., Anno M., Nagashima K.,
 RA Nagashima T., Ikeda S.-I., Tsuji S., Kanazawa I.;
 RT "SCA17, a novel autosomal dominant cerebellar ataxia caused by an
 RT expanded polyglutamine in TATA-binding protein.";
 RL Hum. Mol. Genet. 10:1441-1448(2001).

RN [13]
 RP POLYMORPHISM OF POLY-GLN REGION.
 RX MEDLINE=21937712; PubMed=11939898;
 RA Silveira I., Miranda C., Guimaraes L., Moreira M.-C., Alonso I.,
 RA Mendonca P., Ferro A., Pinto-Basto J., Coelho J., Ferreira F.,
 RA Poirier J., Parreira E., Vale J., Januario C., Barbot C., Tuna A.,
 RA Barros J., Koide R., Tsuji S., Holmes S.E., Margolis R.L., Jardim L.,
 RA Pandolfo M., Coutinho P., Sequeiros J.;
 RT "Trinucleotide repeats in 202 families with ataxia: a small expanded
 RT (CAG)_n allele at the SCA17 locus.";
 RL Arch. Neurol. 59:623-629(2002).

CC -!- FUNCTION: General transcription factor that functions at the
 CC core of the DNA-binding multiprotein factor TFIID. Binding of
 CC TFIID to the TATA box is the initial transcriptional step of the
 CC pre-initiation complex (PIC), playing a role in the activation of
 CC eukaryotic genes transcribed by RNA polymerase II.
 CC -!- SUBUNIT: Belongs to the TFIID complex together with the TBP-
 CC associated factors (TAFs). Binds DNA as monomer. Interacts with
 CC TAFs, TFIIA, TFIIB, NCOA6, DRAP1 and DR1.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- POLYMORPHISM: The poly-Gln region of TBP is highly polymorphic (25
 CC to 42 repeats) in normal individuals and is expanded to about 47-
 CC 63 repeats in SCA17 patients. Longer expansions may result in
 CC earlier onset and more severe clinical manifestations of the
 CC disease.
 CC -!- DISEASE: Defects in TBP are the cause of spinocerebellar ataxia
 CC type 17 (SCA17) [MIM:607136]. SCA17 is a rare autosomal dominant
 CC neurodegenerative disease, characterized by gait ataxia and
 CC dementia, progressing over several decades to include
 CC bradykinesia, dysmetria, dysdiadochokinesis, hyperreflexia and
 CC paucity of movement.
 CC -!- SIMILARITY: Belongs to the TBP family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M55654; AAA36731.1; -.
 DR EMBL; M34960; AAC03409.1; -.
 DR EMBL; X54993; CAA38736.1; -.
 DR EMBL; AL031259; CAA20286.1; -.
 DR PIR; A34830; TWHU2D.
 DR PDB; 1CDW; 23-DEC-96.
 DR PDB; 1C9B; 10-JAN-00.
 DR PDB; 1JFI; 11-JUL-01.
 DR PDB; 1TGH; 01-AUG-96.
 DR TRANSFAC; T00794; -.
 DR Genew; HGNC:11588; TBP.
 DR MIM; 600075; -.
 DR MIM; 607136; -.
 DR GO; GO:0005669; C:transcription factor TFIID complex; TAS.
 DR GO; GO:0016251; F:general RNA polymerase II transcription fac. . .; TAS.
 DR GO; GO:0006367; P:transcription initiation from Pol II promoter; TAS.
 DR InterPro; IPR000814; TFIID.
 DR Pfam; PF00352; TBP; 2.
 DR PRINTS; PR00686; TIFACTORIID.
 DR PROSITE; PS00351; TFIID; 2.
 KW Transcription; Nuclear protein; DNA-binding; Repeat; Polymorphism;
 KW Triplet repeat expansion; Disease mutation; 3D-structure.

FT	REPEAT	165	241	1.
FT	REPEAT	255	332	2.
FT	DOMAIN	55	95	POLY-GLN.
FT	VARIANT	92	95	Missing.
FT				/FTId=VAR_016987.
FT	CONFLICT	187	187	A -> R (IN REF. 2).
FT	STRAND	164	173	
FT	HELIX	180	186	
FT	TURN	188	189	
FT	STRAND	190	192	
FT	TURN	194	196	
FT	STRAND	199	205	
FT	TURN	206	207	
FT	STRAND	208	214	
FT	TURN	215	216	
FT	STRAND	218	222	
FT	HELIX	227	244	
FT	TURN	245	245	
FT	STRAND	251	263	
FT	STRAND	268	268	
FT	HELIX	270	276	
FT	TURN	277	280	
FT	STRAND	281	282	
FT	TURN	285	287	
FT	STRAND	291	295	
FT	TURN	296	299	
FT	STRAND	300	304	
FT	TURN	306	307	
FT	STRAND	309	313	
FT	HELIX	318	333	
FT	TURN	334	335	
FT	STRAND	336	336	

Query Match 56.9%; Score 211.5; DB 1; Length 339;
 Best Local Similarity 67.7%; Pred. No. 1e-10;
 Matches 44; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

QY 1 LVPRGSMATLEKLMKAFESLKSFOQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 60
 :| | : | : : || :||:|||||||||||||||||||||||||||||||||
 Db 31 MMPYGTGLTPQPIQNT-NSLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 89
 QY 61 QQQQQ 65
 |||||
 Db 90 QQQQQ 94

RESULT 5

FXP2_MOUSE

ID FXP2_MOUSE STANDARD; PRT; 714 AA.
 AC P58463;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Forkhead box protein P2.
 GN FOXP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Lung;
 RX MEDLINE=21347947; PubMed=11358962;
 RA Shu W., Yang H., Zhang L., Lu M.M., Morrissey E.E.;
 RT "Characterization of a new subfamily of winged-helix/forkhead (Fox)
 RT genes that are expressed in the lung and act as transcriptional
 RT repressors.";
 RL J. Biol. Chem. 276:27488-27497(2001).
 CC -!- FUNCTION: Transcriptional repressor that play an important role in
 CC the specification and differentiation of lung epithelium. May play
 CC important roles in developing neural, gastrointestinal and
 CC cardiovascular tissues.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: Highest expression in lung. Lower expression
 CC in spleen, skeletal muscle, brain, kidney and small intestine.
 CC -!- DEVELOPMENTAL STAGE: Expressed in developing lung (only distal
 CC epithelium), neural, intestinal and cardiovascular tissues.
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF339106; AAK69651.1; -.

DR MGD; MGI:2148705; Foxp2.
 DR GO; GO:0016564; F:transcriptional repressor activity; IDA.
 DR GO; GO:0016481; P:negative regulation of transcription; IDA.
 DR InterPro; IPR001766; TF_Fork_head.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
 DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
 DR PROSITE; PS50039; FORK_HEAD_3; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein.
 FT ZN_FING 345 370 C2H2-TYPE.
 FT DNA_BIND 503 593 FORK-HEAD.
 FT DOMAIN 53 56 POLY-GLN.
 FT DOMAIN 123 126 POLY-GLN.
 FT DOMAIN 131 136 POLY-GLN.
 FT DOMAIN 152 191 POLY-GLN.
 FT DOMAIN 200 208 POLY-GLN.
 FT DOMAIN 222 230 POLY-GLN.
 SQ SEQUENCE 714 AA; 79820 MW; BCDFB80E28398609 CRC64;

Query Match 56.7%; Score 211; DB 1; Length 714;
 Best Local Similarity 73.3%; Pred. No. 2e-10;
 Matches 44; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

Qy 10 LEKLMKAFESLKSFQQQLQPG 69
 |:: | : | |||||
 Db 135 LQEFYKKQEQQLHLQLLQQQHPG 194

RESULT 6

FXP2_HUMAN

ID FXP2_HUMAN STANDARD; PRT; 715 AA.
 AC O15409; Q8NOW2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Forkhead box protein P2 (CAG repeat protein 44) (Trinucleotide repeat-
 DE containing gene 10 protein).
 GN FOXP2 OR CAGH44 OR TNRC10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT SPCH1 HIS-553.
 RX MEDLINE=21470412; PubMed=11586359;
 RA Lai C.S.L., Fisher S.E., Hurst J.A., Vargha-Khadem F., Monaco A.P.;
 RT "A forkhead-domain gene is mutated in a severe speech and language
 RT disorder."
 RL Nature 413:519-523(2001).

RN [2]
 RP SEQUENCE OF 1-304 FROM N.A.
 RC TISSUE=Brain cortex;
 RX MEDLINE=97369492; PubMed=9225980;
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,
 RA Breschel T.S., Stine O.C., Callahan C., Mcinnis M.G., Ross C.A.;
 RT "cDNAs with long CAG trinucleotide repeats from human brain."
 RL Hum. Genet. 100:114-122(1997).
 RN [3]
 RP SEQUENCE OF 1-86 FROM N.A.
 RA Minx P., Hinds K., Sutterer C., Becker M., Ozersky P.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 113-329 FROM N.A.
 RX MEDLINE=22179809; PubMed=12192408;
 RA Enard W., Przeworski M., Fisher S.E., Lai C.S.L., Wiebe V., Kitano T.,
 RA Monaco A.P., Paabo S.;
 RT "Molecular evolution of FOXP2, a gene involved in speech and
 RT language."
 RL Nature 418:869-872(2002).
 CC -!- FUNCTION: Transcriptional repressor that plays an important role
 CC in the specification and differentiation of lung epithelium. May
 CC play important roles in developing neural, gastrointestinal and
 CC cardiovascular tissues. Involved in neural mechanisms mediating
 CC the development of speech and language.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1; Synonyms=I;
 CC IsoId=O15409-1; Sequence=Displayed;
 CC Name=2; Synonyms=II;
 CC IsoId=O15409-3; Sequence=Not described;
 CC Name=3; Synonyms=III, IV;
 CC IsoId=O15409-2; Sequence=VSP_001558;
 CC -!- TISSUE SPECIFICITY: Expressed at high levels in embryonic and
 CC adult lung.
 CC -!- DISEASE: Defects in FOXP2 are the cause of speech-language
 CC disorder 1 (SPCH1) [MIM:602081]; also known as autosomal dominant
 CC speech and language disorder with orofacial dyspraxia. Affected
 CC individuals have a severe impairment in the selection and
 CC sequencing of fine orofacial movements, which are necessary for
 CC articulation. They also show deficits in several facets of
 CC language processing (such as the ability to break up words into
 CC their constituent phoneme) and grammatical skills.
 CC -!- DISEASE: Disruption of FOXP2 by a chromosomal translocation
 CC t(5;7)(q22;q31.2) is the cause of severe speech and language
 CC impairment.
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; AF337817; AAL10762.1; -.
DR EMBL; U80741; AAB91439.1; -.
DR EMBL; AC003992; -; NOT_ANNOTATED_CDS.
DR EMBL; AF515031; AAN03389.1; -.
DR EMBL; AF515032; AAN03390.1; -.
DR EMBL; AF515033; AAN03391.1; -.
DR EMBL; AF515034; AAN03392.1; -.
DR EMBL; AF515035; AAN03393.1; -.
DR EMBL; AF515036; AAN03394.1; -.
DR EMBL; AF515037; AAN03395.1; -.
DR EMBL; AF515038; AAN03396.1; -.
DR EMBL; AF515039; AAN03397.1; -.
DR EMBL; AF515040; AAN03398.1; -.
DR EMBL; AF515041; AAN03399.1; -.
DR EMBL; AF515042; AAN03400.1; -.
DR EMBL; AF515043; AAN03401.1; -.
DR EMBL; AF515044; AAN03402.1; -.
DR EMBL; AF515045; AAN03403.1; -.
DR EMBL; AF515046; AAN03404.1; -.
DR EMBL; AF515047; AAN03405.1; -.
DR EMBL; AF515048; AAN03406.1; -.
DR EMBL; AF515049; AAN03407.1; -.
DR EMBL; AF515050; AAN03408.1; -.
DR Genew; HGNC:13875; FOXP2.
DR MIM; 605317; -.
DR MIM; 602081; -.
DR InterPro; IPR001766; TF_Fork_head.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Chromosomal translocation; Disease mutation;
KW Alternative splicing.
FT ZN_FING 346 371 C2H2-TYPE.
FT DNA_BIND 504 594 FORK-HEAD.
FT DOMAIN 53 56 POLY-GLN.
FT DOMAIN 123 126 POLY-GLN.
FT DOMAIN 131 136 POLY-GLN.
FT DOMAIN 152 191 POLY-GLN.
FT DOMAIN 200 209 POLY-GLN.
FT DOMAIN 223 231 POLY-GLN.
FT VARSPLIC 1 92 Missing (in isoform 3).
FT /FTid=VSP_001558.
FT VARIANT 553 553 R -> H (in SPCH1).
FT /FTid=VAR_012278.
FT CONFLICT 134 134 Q -> H (IN REF. 2).
FT CONFLICT 290 304 DLTTNNSSSTTSSNT -> EEFPVQGPAAVCAGL (IN
FT REF. 2).

```

SQ SEQUENCE 715 AA; 79919 MW; 4F9FBDB6D90516E0 CRC64;

```

QY      10 LEKLMKAFESLKSFOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOLQPG 69
        |:: | :      | |||||
Db      135 LQEFYKKKQQLHLQLLQOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOHPG 194

```

TAGB DICDI

Db 1814 EQQEQQEQQQQQQQEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQNDQP 1864

RESULT 8

HCN1_MOUSE

ID HCN1_MOUSE STANDARD; PRT; 910 AA.
AC 088704; 054899; Q9D613;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated
DE channel 1 (Brain cyclic nucleotide gated channel 1) (BCNG-1)
DE (Hyperpolarization-activated cation channel 2) (HAC-2).
GN HCN1 OR BCNG1 OR HAC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND N-GLYCOSYLATION.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=98070835; PubMed=9405696;
RA Santoro B., Grant S.G.N., Bartsch D., Kandel E.R.;
RT "Interactive cloning with the SH3 domain of N-src identifies a new
RT brain specific ion channel protein, with homology to eag and cyclic
RT nucleotide-gated channels.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:14815-14820(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=98295993; PubMed=9634236;
RA Ludwig A., Zong X., Jeglitsch M., Hofmann F., Biel M.;
RT "A family of hyperpolarization-activated cation channels.";
RL Nature 393:587-591(1998).
RN [3]
RP SEQUENCE OF 377-910 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [4]
 RP FUNCTION, AND REGULATION BY CAMP.
 RX MEDLINE=98292171; PubMed=9630217;
 RA Santoro B., Liu D.T., Yao H., Bartsch D., Kandel E.R.,
 RA Siegelbaum S.A., Tibbs G.R.;
 RT "Identification of a gene encoding a hyperpolarization-activated
 RT pacemaker channel of brain.";
 RL Cell 93:717-729(1998).
 RN [5]
 RP INTERACTION WITH KCNE2.
 RX MEDLINE=21313430; PubMed=11420311;
 RA Yu H., Wu J., Potapova I., Wymore R.T., Holmes B., Zuckerman J.,
 RA Pan Z., Wang H., Shi W., Robinson R.B., El-Maghrabi M.R., Benjamin W.,
 RA Dixon J.E., McKinnon D., Cohen I.S., Wymore R.;
 RT "MinK-related peptide 1: A beta subunit for the HCN ion channel
 RT subunit family enhances expression and speeds activation.";
 RL Circ. Res. 88:E84-E87(2001).
 RN [6]
 RP REGULATION BY CAMP.
 RX MEDLINE=21351681; PubMed=11459060;
 RA Wainger B.J., DeGennaro M., Santoro B., Siegelbaum S.A., Tibbs G.R.;
 RT "Molecular mechanism of cAMP modulation of HCN pacemaker channels.";
 RL Nature 411:805-810(2001).
 RN [7]
 RP FUNCTION, AND TISSUE SPECIFICITY.
 RX MEDLINE=21530492; PubMed=11675786;
 RA Stevens D.R., Seifert R., Bufe B., Mueller F., Kremmer E., Gauss R.,
 RA Meyerhof W., Kaupp U.B., Lindemann B.;
 RT "Hyperpolarization-activated channels HCN1 and HCN4 mediate responses
 RT to sour stimuli.";
 RL Nature 413:631-635(2001).
 RN [8]
 RP INTERACTION WITH HCN2, AND MUTAGENESIS OF GLY-349; TYR-350 AND
 RP GLY-351.
 RX MEDLINE=22083667; PubMed=12089064;
 RA Xue T., Marban E., Li R.A.;
 RT "Dominant-negative suppression of HCN1- and HCN2-encoded pacemaker
 RT currents by an engineered HCN1 construct: insights into
 RT structure-function relationships and multimerization.";
 RL Circ. Res. 90:1267-1273(2002).
 RN [9]
 RP OLIGOMERIZATION VIA N-TERMINAL DOMAIN.
 RX MEDLINE=22162449; PubMed=12034718;
 RA Proenza C., Tran N., Angoli D., Zahynacz K., Balcar P., Accili E.A.;
 RT "Different roles for the cyclic nucleotide binding domain and amino
 RT terminus in assembly and expression of hyperpolarization-activated,
 RT cyclic nucleotide-gated channels.";
 RL J. Biol. Chem. 277:29634-29642(2002).
 RN [10]
 RP MUTAGENESIS OF CYS-303 AND CYS-318.
 RX MEDLINE=22336443; PubMed=12351622;
 RA Xue T., Li R.A.;
 RT "An external determinant in the S5-P linker of the pacemaker (HCN)
 RT channel identified by sulfhydryl modification.";
 RL J. Biol. Chem. 277:46233-46242(2002).

CC -!- FUNCTION: Hyperpolarization-activated ion channel exhibiting weak
 CC selectivity for potassium over sodium ions. Contributes to the
 CC native pacemaker currents in heart (If) and in neurons (Ih).
 CC Activated by cAMP, and at 10-100 times higher concentrations, also
 CC by cGMP. May mediate responses to sour stimuli.
 CC -!- SUBUNIT: The potassium channel is probably composed of a homo- or
 CC heterotetrameric complex of pore-forming subunits. Heteromultimer
 CC with HCN2. Interacts with KCNE2. Interacts with the SH3 domain of
 CC CSK.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Predominantly expressed in brain. Highly
 CC expressed in apical dendrites of pyramidal neurons in the cortex,
 CC in the layer corresponding to the stratum lacunosum-moleculare in
 CC the hippocampus and in axons of basket cells in the cerebellum.
 CC Expressed in a subset of elongated cells in taste buds.
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
 CC characterized by a series of positively charged amino acids at
 CC every third position.
 CC -!- PTM: N-glycosylated.
 CC -!- MISCELLANEOUS: Inhibited by extracellular cesium ions.
 CC -!- SIMILARITY: Belongs to the potassium channel family. HCN
 CC subfamily.
 CC -!- SIMILARITY: Contains 1 cyclic nucleotide-binding domain.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
 CC frameshift in position 381.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF028737; AAC53518.1; -.
 DR EMBL; AJ225123; CAA12407.1; -.
 DR EMBL; AK014722; BAB29519.1; ALT_FRAME.
 DR MGD; MGI:1096392; Hcn1.

DR InterPro; IPR000595; cNMP_binding.
 DR InterPro; IPR005821; Ion_trans.
 DR InterPro; IPR001622; K+channel_pore.
 DR InterPro; IPR005820; M+channel_nlg.
 DR Pfam; PF00027; cNMP_binding; 1.
 DR Pfam; PF00520; ion_trans; 1.
 DR SMART; SM00100; cNMP; 1.
 DR PROSITE; PS00888; CNMP_BINDING_1; 1.
 DR PROSITE; PS00889; CNMP_BINDING_2; FALSE_NEG.
 DR PROSITE; PS50042; CNMP_BINDING_3; 1.

KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
 KW Potassium channel; Potassium; Potassium transport; Sodium transport;
 KW cAMP; cAMP-binding; Transmembrane; Glycoprotein; Sodium channel.

FT	DOMAIN	1	135	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	136	156	SEGMENT S1 (POTENTIAL).
FT	TRANSMEM	163	183	SEGMENT S2 (POTENTIAL).
FT	DOMAIN	184	208	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	209	229	SEGMENT S3 (POTENTIAL).
FT	TRANSMEM	238	258	SEGMENT S4 (POTENTIAL).

FT	DOMAIN	259	289	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	290	310	SEGMENT S5 (POTENTIAL).
FT	TRANSMEM	334	355	SEGMENT H5 (PORE-FORMING) (POTENTIAL).
FT	TRANSMEM	361	381	SEGMENT S6 (POTENTIAL).
FT	DOMAIN	382	910	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	78	129	INVOLVED IN SUBUNIT ASSEMBLY (BY
FT				SIMILARITY).
FT	NP_BIND	464	581	CAMP.
FT	DOMAIN	1	81	GLY-RICH.
FT	DOMAIN	715	777	GLN-RICH.
FT	DOMAIN	878	884	POLY-PRO.
FT	CARBOHYD	327	327	N-LINKED (GLCNAC. . .) (PROBABLE).
FT	MUTAGEN	303	303	C->S: ABOLISHES CONDUCTIVITY.
FT	MUTAGEN	318	318	C->S: ABOLISHES SENSITIVITY TO SULFHYDRIL
FT				MODIFICATION.
FT	MUTAGEN	349	349	G->A: ABOLISHES CONDUCTIVITY; WHEN
FT				ASSOCIATED WITH A-350 AND A-351.
FT	MUTAGEN	350	350	Y->A: ABOLISHES CONDUCTIVITY; WHEN
FT				ASSOCIATED WITH A-349 AND A-351.
FT	MUTAGEN	351	351	G->A: ABOLISHES CONDUCTIVITY; WHEN
FT				ASSOCIATED WITH A-349 AND A-350.
FT	CONFLICT	42	42	G -> R (IN REF. 1).
FT	CONFLICT	394	394	R -> S (IN REF. 3).
SQ	SEQUENCE	910 AA;	102432 MW;	56FD5F328DD972E9 CRC64;

```

ID      HD HUMAN          STANDARD;          PRT;   3144 AA.
AC      P42858;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Huntingtin (Huntington's disease protein) (HD protein).
GN      HD OR IT15.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Retina;
RX      MEDLINE=93208892; PubMed=8458085;
RA      Macdonald M., Ambrose C.M., Duyao M.P., Myers R.H., Lin C.S.,
RA      Srinidhi J., Barnes G., Taylor S.A., James M., Groot N., McFarlane H.,
RA      Jenkins B., Anderson M.A., Wexler N.S., Gusella J.F., Bates G.P.,
RA      Baxendale S., Hummerich H., Kirby S., North M., Youngman S., Mott R.,
RA      Zehetner G., Sedlacek Z., Poustka A., Frischauf A.-M., Lehrach H.,
RA      Buckler A.J., Church D., Doucette-Stamm L., O'Donovan M.C.,

```


RA Riba-Ramirez L., Shah M., Stanton V.P., Strobel S.A., Draths K.M.,
 RA Wales J.L., Dervan P., Housman D.E., Altherr M., Shiang R.,
 RA Thompson L., Fielder T., Wasmuth J.J., Tagle D., Valdes J., Elmer L.,
 RA Allard M., Castilla L., Swaroop M., Blanchard K., Collins F.S.,
 RA Snell R., Holloway T., Gillespie K., Datson N., Shaw S., Harper P.S.;
 RT "A novel gene containing a trinucleotide repeat that is expanded and
 RT unstable on Huntington's disease chromosomes. The Huntington's
 RT Disease Collaborative Research Group.";
 RL Cell 72:971-983(1993).
 RN [2]
 RP SEQUENCE OF 1-90 FROM N.A.
 RX MEDLINE=95278941; PubMed=7759106;
 RA Lin B., Nasir J., Kalchman M.A., McDonald H., Zeisler J.,
 RA Goldberg Y.P., Hayden M.R.;
 RT "Structural analysis of the 5' region of mouse and human Huntington
 RT disease genes reveals conservation of putative promoter region and
 RT di- and trinucleotide polymorphisms.";
 RL Genomics 25:707-715(1995).
 RN [3]
 RP SEQUENCE OF 1-205 FROM N.A.
 RX MEDLINE=94255787; PubMed=8197474;
 RA Ambrose C.M., Duyao M.P., Barnes G., Bates G.P., Lin C.S.,
 RA Srinidhi J., Baxendale S., Hummerich H., Lehrach H., Altherr M.,
 RA Wasmuth J., Buckler A., Church D., Housman D., Berks M., Micklem G.,
 RA Durbin R., Dodge A., Read A., Gusella J.F., Macdonald M.E.;
 RT "Structure and expression of the Huntington's disease gene: evidence
 RT against simple inactivation due to an expanded CAG repeat.";
 RL Somat. Cell Mol. Genet. 20:27-38(1994).
 RN [4]
 RP SEQUENCE OF 1-117 FROM N.A.
 RA Matthews P.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SEQUENCE OF 119-934 FROM N.A.
 RA Lloyd C.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
 RN [6]
 RP SEQUENCE OF 1212-1290 FROM N.A.
 RA Mungall A., Odell C.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
 RN [7]
 RP SEQUENCE OF 1291-1860 FROM N.A.
 RA Mungall A.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
 RN [8]
 RP SEQUENCE OF 1862-2820 FROM N.A.
 RA Buck D.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
 RN [9]
 RP SEQUENCE OF 2563-3144 FROM N.A.
 RC TISSUE=Brain, Caudate, Frontal cortex, Muscle, and Retina;
 RX MEDLINE=94093536; PubMed=7903579;
 RA Lin B., Rommens J.M., Graham R.K., Kalchman M., Macdonald H.,
 RA Nasir J., Delaney A., Goldberg Y.P., Hayden M.R.;
 RT "Differential 3' polyadenylation of the Huntington disease gene
 RT results in two mRNA species with variable tissue expression.";
 RL Hum. Mol. Genet. 2:1541-1545(1993).

RN [10]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=95375771; PubMed=7647777;
 RA Trottier Y., Devys D., Imbert G., Saudou F., An I., Lutz Y., Weber C.,
 RA Agid Y., Hirsch E.C., Mandel J.-L.;
 RT "Cellular localization of the Huntington's disease protein and
 RT discrimination of the normal and mutated form.";
 RL Nat. Genet. 10:104-110(1995).
 RN [11]
 RP CLEAVAGE BY APOPAIN.
 RX MEDLINE=96331285; PubMed=8696339;
 RA Goldberg Y.P., Nicholson D.W., Rasper D.M., Kalchman M.A., Koide H.B.,
 RA Graham R.K., Bromm M., Kazemi-Esfarjani P., Thornberry N.A.,
 RA Vaillancourt J.P., Hayden M.R.;
 RT "Cleavage of huntingtin by apopain, a proapoptotic cysteine protease,
 RT is modulated by the polyglutamine tract.";
 RL Nat. Genet. 13:442-449(1996).
 RN [12]
 RP INTERACTION WITH FBNP3.
 RX MEDLINE=98367036; PubMed=9700202;
 RA Faber P.W., Barnes G.T., Srinidhi J., Chen J., Gusella J.F.,
 RA MacDonald M.E.;
 RT "Huntingtin interacts with a family of WW domain proteins.";
 RL Hum. Mol. Genet. 7:1463-1474(1998).
 CC -!- FUNCTION: May play a role in microtubule-mediated transport or
 CC vesicle function.
 CC -!- SUBUNIT: Binds SH3GLB1 (By similarity). Interacts through its N-
 CC terminus with FBNP3.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: Widely expressed with the highest level of
 CC expression in the brain (nerve fibers, varicosities, and nerve
 CC endings). In the brain, the regions where it can be mainly found
 CC are the cerebellar cortex, the neocortex, the striatum, and the
 CC hippocampal formation.
 CC -!- PTM: Cleaved by apopain downstream of the polyglutamine stretch.
 CC The resulting amino-terminal fragment is cytotoxic and provokes
 CC apoptosis.
 CC -!- POLYMORPHISM: The poly-Gln region of HD is highly polymorphic (10
 CC to 35 repeats) in the normal population and is expanded to about
 CC 36-120 repeats in hd patients. The repeat length usually increases
 CC in successive generations, but contracts also on occasion. The
 CC longer expansions result in earlier onset and more severe clinical
 CC manifestations of the disease. The adjacent poly-pro region is
 CC also polymorphic and varies between 7-12 residues. Polyglutamine
 CC expansion leads to elevated susceptibility to apopain cleavage and
 CC likely result in accelerated neuronal apoptosis.
 CC -!- DISEASE: DEFECTS IN HD ARE THE CAUSE OF HUNTINGTON'S DISEASE, AN
 CC AUTOSOMAL DOMINANT NEURODEGENERATIVE DISORDER CHARACTERIZED BY
 CC INVOLUNTARY MOVEMENTS (CHOREA), GENERAL MOTOR IMPAIRMENT,
 CC PSYCHIATRIC DISORDERS AND DEMENTIA. ONSET OF THE DISEASE OCCURS
 CC USUALLY IN THE THIRD OR FOURTH DECADE OF LIFE AND SYMPTOMS
 CC PROGRESSIVELY WORSEN LEADING TO DEATH IN 10 TO 20 YEARS. IT
 CC AFFECTS 1 IN 10,000 INDIVIDUALS OF EUROPEAN ORIGIN. NEUROPATHOLOGY
 CC OF HUNTINGTON'S DISEASE DISPLAYS A DISTINCTIVE PATTERN WITH LOSS
 CC OF NEURONS, SPECIALLY IN THE CAUDATE AND PUTAMEN (STRIATUM).
 CC -!- SIMILARITY: Contains 10 HEAT repeats.
 CC -!- SIMILARITY: Belongs to the huntingtin family.

```

CC  -!- DATABASE: NAME=HotMolecBase; NOTE=HD entry;
CC
WWW="http://bioinformatics.weizmann.ac.il/hotmolecbase/entries/hunti.htm".
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L12392; AAB38240.1; -.
DR  EMBL; L34020; -; NOT_ANNOTATED_CDS.
DR  EMBL; L27350; -; NOT_ANNOTATED_CDS.
DR  EMBL; L27351; -; NOT_ANNOTATED_CDS.
DR  EMBL; L27352; -; NOT_ANNOTATED_CDS.
DR  EMBL; L27353; -; NOT_ANNOTATED_CDS.
DR  EMBL; L27354; -; NOT_ANNOTATED_CDS.
DR  EMBL; Z68756; -; NOT_ANNOTATED_CDS.
DR  EMBL; Z49155; CAA89025.1; -.
DR  EMBL; Z49208; -; NOT_ANNOTATED_CDS.
DR  EMBL; Z69649; -; NOT_ANNOTATED_CDS.
DR  EMBL; Z49154; CAA89024.1; -.
DR  EMBL; Z49769; CAA89839.1; -.
DR  EMBL; L20431; AAA52702.1; -.
DR  PIR; A46068; A46068.
DR  Genew; HGNC:4851; HD.
DR  MIM; 143100; -.
DR  GO; GO:0005737; C:cytoplasm; TAS.
DR  GO; GO:0005634; C:nucleus; TAS.
DR  GO; GO:0005625; C:soluble fraction; TAS.
DR  GO; GO:0008017; F:microtubule binding; TAS.
DR  GO; GO:0005515; F:protein binding; IPI.
DR  GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR  GO; GO:0005215; F:transporter activity; TAS.
DR  GO; GO:0007610; P:behavior; TAS.
DR  GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR  GO; GO:0006917; P:induction of apoptosis; TAS.
DR  GO; GO:0009405; P:pathogenesis; TAS.
DR  InterPro; IPR000091; Huntingtin.
DR  Pfam; PF03541; Huntingtin; 1.
DR  PRINTS; PR00375; HUNTINGTIN.
KW  Repeat; Disease mutation; Polymorphism; Triplet repeat expansion;
KW  Apoptosis.
FT  DOMAIN      205      329      HEAT REPEATS DOMAIN 1.
FT  DOMAIN      745      942      HEAT REPEATS DOMAIN 2.
FT  DOMAIN     1534     1575      HEAT REPEATS DOMAIN 3.
FT  DOMAIN       18       40      POLY-GLN.
FT  DOMAIN       41       51      POLY-PRO.
FT  DOMAIN       65       80      POLY-PRO.
FT  DOMAIN     1439     1442      POLY-THR.
FT  DOMAIN     2343     2347      POLY-GLU.
FT  DOMAIN     2640     2645      POLY-GLU.
FT  SITE        513      514      CLEAVAGE (BY APOPAIN) (POTENTIAL).
FT  SITE        530      531      CLEAVAGE (BY APOPAIN) (POTENTIAL).
FT  SITE        552      553      CLEAVAGE (BY APOPAIN) (POTENTIAL).

```

FT SITE 589 590 CLEAVAGE (BY APOPAIN) (POTENTIAL).
 FT VARIANT 38 40 Missing.
 FT /FTId=VAR_005268.
 FT CONFLICT 2788 2788 V -> I (IN REF. 10).
 SQ SEQUENCE 3144 AA; 347855 MW; 9D1BA8528929908F CRC64;

Query Match 54.8%; Score 204; DB 1; Length 3144;
 Best Local Similarity 72.6%; Pred. No. 2.5e-09;
 Matches 45; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 7 MATLEKLMKAFESLKSFQQQL 66
 |||||
 Db 1 MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPQLPQPPPPQA 60
 Qy 67 QP 68
 ||
 Db 61 QP 62

RESULT 10

T230_HUMAN

ID T230_HUMAN STANDARD; PRT; 2212 AA.
 AC Q93074; O15410; O75557; Q9UHV6; Q9UND7;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Thyroid hormone receptor-associated protein complex 230 kDa component
 DE (Trap230) (Activator-recruited cofactor 240 kDa component) (ARC240)
 DE (CAG repeat protein 45) (OPA-containing protein) (Trinucleotide repeat
 DE containing 11).
 GN TNRC11 OR TRAP230 OR ARC240 OR CAGH45 OR HOPA OR KIAA0192.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99214851; PubMed=10198638;
 RA Ito M., Yuan C.-X., Malik S., Gu W., Fondell J.D., Yamamura S.,
 RA Fu Z.-Y., Zhang X., Qin J., Roeder R.G.;
 RT "Identity between TRAP and SMCC complexes indicates novel pathways for
 RT the function of nuclear receptors and diverse mammalian activators.";
 RL Mol. Cell 3:361-370(1999).
 RN [2]
 RP SEQUENCE OF 89-2212 FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96281124; PubMed=8724849;
 RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. V.
 RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 3:17-24(1996).
 RN [3]
 RP SEQUENCE OF 189-2212 FROM N.A.
 RX MEDLINE=98368120; PubMed=9702738;
 RA Philibert R.A., King B.H., Cook E.H., Lee Y.-H., Stubblefield B.,
 RA Damschroder-Williams P., Dea C., Palotie A., Tengstrom C.,

RA Martin B.M., Ginns E.I.;
 RT "Association of an X-chromosome dodecamer insertional variant allele
 RT with mental retardation.";
 RL Mol. Psych. 3:303-309(1998).
 RN [4]
 RP SEQUENCE OF 189-2212 FROM N.A.
 RX MEDLINE=99408253; PubMed=10480376;
 RA Philibert R.A., Winfield S.L., Damschroder-Williams P., Tengstrom C.,
 RA Martin B.M., Ginns E.I.;
 RT "The genomic structure and developmental expression patterns of the
 RT human OPA-containing gene (HOPA).";
 RL Hum. Genet. 105:174-178(1999).
 RN [5]
 RP SEQUENCE OF 1564-2212 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97369492; PubMed=9225980;
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
 RT "cDNAs with long CAG trinucleotide repeats from human brain.";
 RL Hum. Genet. 100:114-122(1997).
 RN [6]
 RP IDENTIFICATION IN ARC COMPLEX, AND SEQUENCE OF 1709-1717 AND
 RP 1806-1817.
 RX MEDLINE=99249346; PubMed=10235267;
 RA Naeaer A.M., Beurang P.A., Zhou S., Abraham S., Solomon W.B.,
 RA Tjian R.;
 RT "Composite co-activator ARC mediates chromatin-directed
 RT transcriptional activation.";
 RL Nature 398:828-832(1999).
 CC -!- FUNCTION: Plays a role in transcriptional coactivation.
 CC -!- SUBUNIT: Subunit of the large multiprotein complexes TRAP and
 CC ARC/DRIP.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF117755; AAD22033.1; -.
 DR EMBL; D83783; BAA12112.1; -.
 DR EMBL; AF071309; AAC83163.1; -.
 DR EMBL; AF132033; AAD44162.1; -.
 DR EMBL; U80742; AAB91440.1; -.
 DR Genew; HGNC:11957; TNRC11.
 DR MIM; 300188; -.
 DR GO; GO:0000119; C:mediator complex; IDA.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0030374; F:ligand-dependent nuclear receptor transcrip. . .; NAS.
 DR GO; GO:0004872; F:receptor activity; IDA.
 DR GO; GO:0016455; F:RNA polymerase II transcription mediator ac. . .; IDA.
 DR GO; GO:0046966; F:thyroid hormone receptor binding; IDA.
 DR GO; GO:0016563; F:transcriptional activator activity; IDA.

RT "A mutant *Drosophila* homolog of mammalian Clock disrupts circadian
 RT rhythms and transcription of period and timeless.";
 RL Cell 93:791-804(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=98414630; PubMed=9742131;
 RA Bae K., Lee C., Sidote D., Chuang K.-Y., Edery I.;
 RT "Circadian regulation of a *Drosophila* homolog of the mammalian clock
 RT gene: PER and TIM function as positive regulators.";
 RL Mol. Cell. Biol. 18:6142-6151(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: Circadian regulator that acts as a transcription factor
 CC and generates a rhythmic output with a period of about 24 hours.
 CC Oscillates in antiphase to the cycling observed for period (PER)
 CC and timeless (TIM). According to Ref.3, reaches peak abundance

CC within several hours of the dark-light transition at ZT0
 CC (zeitgeber 0), whereas Ref.1 describes bimodal oscillating
 CC expression with maximum at ZT5 and ZT23. Clock-cycle heterodimers
 CC activate cycling transcription of PER and TIM by binding to the e-
 CC box (3'-CACGTG-5') present in their promoters. Once induced,
 CC Period and Timeless block Clock's ability to transactivate their
 CC promoters.
 CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
 CC bHLH protein. Forms a heterodimer with Cycle.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- TISSUE SPECIFICITY: Widely expressed. Found in head, body, and
 CC appendage fractions.
 CC -!- DOMAIN: Contains three polyglutamine repeats which could
 CC correspond to the transactivation domain. The length of the
 CC repeats is polymorphic. In the arrhythmic mutant JRK, deletion of
 CC this region leads to the loss of circadian rhythmicity and altered
 CC light response.
 CC -!- POLYMORPHISM: The variability in length of the polyglutamine
 CC stretch is due to polymorphism of this region. Variant B encodes
 CC two conceptual proteins, the first consists only of the bHLH
 CC domain, the other consists of the PAS-1 and all C-terminal
 CC domains. Variant B is expressed weakly at all the times of the
 CC day, and it cycles in phase with the full-length form.
 CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
 CC -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
 CC -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF067207; AAD10630.1; -.
 DR EMBL; AF065133; AAC39101.1; -.
 DR EMBL; AF069997; AAC62234.1; -.
 DR EMBL; AE003557; AAF50516.1; -.
 DR PIR; T13068; T13068.
 DR PIR; T13071; T13071.
 DR FlyBase; FBgn0023076; Clk.
 DR GO; GO:0005634; C:nucleus; NAS.
 DR GO; GO:0003677; F:DNA binding; NAS.
 DR GO; GO:0008062; P:eclosion rhythm; NAS.
 DR GO; GO:0045475; P:locomotor rhythm; NAS.
 DR GO; GO:0045893; P:positive regulation of transcription, DNA-d. . .; IGI.
 DR GO; GO:0045187; P:regulation of sleep; IMP.
 DR GO; GO:0008341; P:response to cocaine (sensu Insecta); NAS.
 DR InterPro; IPR001092; HLH_basic.
 DR InterPro; IPR001067; Nuc_translocat.
 DR InterPro; IPR001610; PAC.
 DR InterPro; IPR000014; PAS_domain.
 DR Pfam; PF00010; HLH; 1.
 DR Pfam; PF00785; PAC; 1.
 DR Pfam; PF00989; PAS; 1.
 DR PRINTS; PR00785; NCTRNSLOCATR.

RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
 RT trithorax.";
 RL Oncogene 15:549-560(1997).
 RN [2]
 RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=22371496; PubMed=12482968;
 RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
 RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
 RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,
 RA Lee J.W.;
 RT "Activating signal cointegrator 2 belongs to a novel steady-state
 RT complex that contains a subset of trithorax group proteins.";
 RL Mol. Cell. Biol. 23:140-149(2003).
 CC -!- FUNCTION: May be involved in transcriptional regulation.
 CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
 CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
 CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
 CC MLL2 and MLL3, and ASH2/ASCL2.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=O14686-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O14686-2; Sequence=VSP_008563, VSP_008559;
 CC Name=3;
 CC IsoId=O14686-3; Sequence=VSP_008560;
 CC -!- TISSUE SPECIFICITY: Expressed in most adult tissues, including a
 CC variety of hematoipoietic cells, with the exception of the liver.
 CC -!- MISCELLANEOUS: This gene mapped to a chromosomal region involved
 CC in duplications and translocations associated with cancer.
 CC -!- SIMILARITY: Belongs to the transcription factor trithorax family.
 CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.
 CC -!- SIMILARITY: Contains 1 post-SET domain.
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
 CC -!- SIMILARITY: Contains 1 SET domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF010403; AAC51734.1; -.
 DR EMBL; AF010404; AAC51735.1; -.
 DR PIR; T03454; T03454.
 DR PIR; T03455; T03455.
 DR Genew; HGNC:7133; MLL2.
 DR MIM; 602113; -.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0003700; F:transcription factor activity; TAS.
 DR GO; GO:0007048; P:oncogenesis; TAS.
 DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
 DR InterPro; IPR003889; FYrich_C.

DR InterPro; IPR003888; FYrich_N.
 DR InterPro; IPR000910; HMG_12_box.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR006118; Recombinase.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00628; PHD; 5.
 DR Pfam; PF00856; SET; 1.
 DR SMART; SM00542; FYRC; 1.
 DR SMART; SM00541; FYRN; 1.
 DR SMART; SM00398; HMG; 1.
 DR SMART; SM00249; PHD; 7.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00184; RING; 3.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS50868; POST_SET; 1.
 DR PROSITE; PS50280; SET; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 5.
 DR PROSITE; PS50016; ZF_PHD_2; 5.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 KW Nuclear protein; Transcription regulation; Coiled coil; Zinc-finger;
 KW Repeat; Alternative splicing; Polymorphism.
 FT ZN_FING 226 276 PHD-TYPE 1.
 FT ZN_FING 229 274 RING-TYPE.
 FT ZN_FING 273 323 PHD-TYPE 2.
 FT ZN_FING 1102 1155 PHD-TYPE 3.
 FT ZN_FING 1152 1202 PHD-TYPE 4.
 FT ZN_FING 1229 1284 PHD-TYPE 5.
 FT DOMAIN 5121 5242 SET.
 FT DOMAIN 5246 5262 POST-SET.
 FT DOMAIN 2397 2436 COILED COIL (POTENTIAL).
 FT DOMAIN 2788 2809 COILED COIL (POTENTIAL).
 FT DOMAIN 2974 3001 COILED COIL (POTENTIAL).
 FT DOMAIN 3286 3342 COILED COIL (POTENTIAL).
 FT DOMAIN 3437 3476 COILED COIL (POTENTIAL).
 FT DOMAIN 3621 3701 COILED COIL (POTENTIAL).
 FT DOMAIN 4265 4287 COILED COIL (POTENTIAL).
 FT DOMAIN 439 668 15 X 5 AA REPEATS OF S/P-P-P-E/P-E/A.
 FT REPEAT 442 446 1.
 FT REPEAT 460 464 2.
 FT REPEAT 469 473 3.
 FT REPEAT 496 500 4.
 FT REPEAT 504 508 5.
 FT REPEAT 521 525 6.
 FT REPEAT 555 559 7.
 FT REPEAT 564 568 8.
 FT REPEAT 573 577 9.
 FT REPEAT 582 586 10.
 FT REPEAT 609 613 11.
 FT REPEAT 618 622 12.
 FT REPEAT 627 631 13.
 FT REPEAT 645 649 14.
 FT REPEAT 663 667 15.
 FT DOMAIN 229 326 CYS-RICH.
 FT DOMAIN 374 922 PRO-RICH.
 FT DOMAIN 1015 1053 ARG-RICH.

FT	DOMAIN	1122	1235	CYS-RICH.
FT	DOMAIN	1832	2351	PRO-RICH.
FT	DOMAIN	2536	2547	GLN-RICH.
FT	DOMAIN	2587	2703	PRO-RICH.
FT	DOMAIN	2986	4000	GLN-RICH.
FT	DOMAIN	3966	4085	PRO-RICH.
FT	DOMAIN	4634	4702	PRO-RICH.
FT	VARSPPLIC	1	305	Missing (in isoform 2).
FT				/FTId=VSP_008563.
FT	VARSPPLIC	306	672	PMEELPAHSWKCKACRVCRACGAGSAELNPNSEWFENYSLC
FT				HRCHKAQGGQTIRSVAEQHTPVCSRFSPPPEPGDTPTDEPDA
FT				LYVACQGQPKGGHVTSMQPKPEGPLQCEAKPLGKAGVQLEP
FT				QLEAPLNEEMPLLLPPEESPLSPPEESPTSPPEASRLSP
FT				PPEELPASPLPEALHLSRPLEESPLSPPEESPLSPPESS
FT				PFSPLLEESPLSPPEESPPSPALETPLSPPEASPLSPPEE
FT				SPLSPPEELPTSPPEASRLSPPEESPMSPPEESPMSP
FT				PPEASRLFPPEESPLSPPEESPLSPPEASRLSPPEES
FT				PMSPPEESPMSPPEEVSRLSPLPVSRLSPPPEESPLS
FT				-> MSPPEESPMSPPEASRLFPPEESPLSPPEESPLS
FT				PPPEASRLSPPEESPMSPPEESPMSPPEEVSRLSPLPV
FT				SRLSPPEESPLSPPEESPTSPPEASRLSPPEESPTSP
FT				PPEDSPASPPEDSLMSLPLEESPLLPLPEEPQLCPRSEGP
FT				HLSRPEEPHLSPRPEEPHLSPQAEAPHLSQPPEEPCLCAV
FT				PEEPHLSPQAEAPHLSQPPEELHLSPQTEEPHLSPVPEEPC
FT				LSPQPEESHLSQPQSEEPCLSPRPEESHLSPELEKPPLSRP
FT				EKPPEEPGQCPAPEELPLFPPEEPGLSPPLLGEPALSEPGE
FT				PPLSPLPEELPLSPSGEPLSLSPQLMPPDPLPPPLSPIITAA
FT				A (in isoform 2).
FT				/FTId=VSP_008559.
FT	VARSPPLIC	1454	1454	E -> EGET (in isoform 3).
FT				/FTId=VSP_008560.
FT	VARIANT	4949	4949	R -> H (in dbSNP:3782356).
FT				/FTId=VAR_017115.
SQ	SEQUENCE	5262 AA;	564171 MW;	26B7C74CAD417E44 CRC64;

Query Match 53.1%; Score 197.5; DB 1; Length 5262;
 Best Local Similarity 66.2%; Pred. No. 1.2e-08;
 Matches 43; Conservative 6; Mismatches 15; Indels 1; Gaps 1;

QY	3	PRGSMATLEKLMKAFESLKS	62
		: :: : : :	
Db	3618	PMGSLQQLQQ-QQQLQQQQQLQQQQQQQLQQQQQLQQQQQLQQQQQLQQQ	3676
QY	63	QQQLQ	67
Db	3677	QQQLQ	3681

RESULT 13

FXP1_MOUSE

ID	FXP1_MOUSE	STANDARD;	PRT;	705 AA.
AC	P58462;			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Forkhead box protein P1 (Forkhead-related transcription factor 1).			
GN	FOXPL.			

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
 RC STRAIN=C57BL/6; TISSUE=Lung;
 RX MEDLINE=21347947; PubMed=11358962;
 RA Shu W., Yang H., Zhang L., Lu M.M., Morrissey E.E.;
 RT "Characterization of a new subfamily of winged-helix/forkhead (Fox)
 RT genes that are expressed in the lung and act as transcriptional
 RT repressors.";
 RL J. Biol. Chem. 276:27488-27497(2001).
 CC -!- FUNCTION: Transcriptional repressor that play an important role in
 CC the specification and differentiation of lung epithelium.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=A;
 CC IsoId=P58462-1; Sequence=Displayed;
 CC Note=Isoform C is produced by alternative initiation at Met-251
 CC of isoform A;
 CC Name=B;
 CC IsoId=P58462-2; Sequence=VSP_001557;
 CC Event=Alternative initiation;
 CC Comment=2 isoforms, A (shown here) and C, are produced by
 CC alternative initiation at Met-1 and Met-251;
 CC -!- TISSUE SPECIFICITY: Highest expression in the lung, brain, and
 CC spleen. Lower expression in heart, skeletal muscle, kidney, small
 CC intestine (isoform C not present) and liver.
 CC -!- DEVELOPMENTAL STAGE: Expressed in developing lung, neural,
 CC intestinal and cardiovascular tissues.
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF339103; AAK69648.1; -.
 DR EMBL; AF339104; AAK69649.1; -.
 DR EMBL; AF339105; AAK69650.1; -.
 DR MGD; MGI:1914004; Foxp1.
 DR GO; GO:0016564; F:transcriptional repressor activity; IDA.
 DR GO; GO:0016481; P:negative regulation of transcription; IDA.
 DR InterPro; IPR001766; TF_Fork_head.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.

CC sequence of the enhancer element of the AFP gene.
 CC -!- SUBUNIT: Interacts with FBNP3.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Contains 4 homeobox domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D26046; BAA05046.1; -.
 DR HSSP; P20263; LOCP.
 DR TRANSFAC; T03881; -.
 DR MGD; MGI:99948; Atbfl.
 DR GO; GO:0005634; C:nucleus; NAS.
 DR GO; GO:0003700; F:transcription factor activity; TAS.
 DR GO; GO:0030182; P:neuron differentiation; TAS.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00046; homeobox; 4.
 DR Pfam; PF00096; zf-C2H2; 18.
 DR ProDom; PD000010; Homeobox; 4.
 DR PROSITE; PS00027; HOMEBOX_1; 2.
 DR PROSITE; PS50071; HOMEBOX_2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
 KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
 KW DNA-binding; Homeobox; Nuclear protein; Repeat.
 FT ZN_FING 79 103 C2H2-TYPE.
 FT ZN_FING 282 305 C2H2-TYPE.
 FT ZN_FING 641 664 C2H2-TYPE.
 FT ZN_FING 672 695 C2H2-TYPE.
 FT ZN_FING 727 751 C2H2-TYPE.
 FT ZN_FING 805 829 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 946 969 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 985 1009 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 1041 1065 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 1089 1113 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 1233 1256 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 1262 1285 C2H2-TYPE.
 FT ZN_FING 1370 1395 C2H2-TYPE.
 FT ZN_FING 1411 1433 C2H2-TYPE.
 FT ZN_FING 1439 1462 C2H2-TYPE.
 FT ZN_FING 1555 1579 C2H2-TYPE.
 FT ZN_FING 1606 1630 C2H2-TYPE.
 FT ZN_FING 1990 2013 C2H2-TYPE.
 FT DNA_BIND 2152 2211 HOMEBOX 1.
 FT DNA_BIND 2249 2308 HOMEBOX 2.
 FT ZN_FING 2335 2358 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 2539 2561 C2H2-TYPE.
 FT DNA_BIND 2650 2709 HOMEBOX 3.
 FT ZN_FING 2720 2743 C2H2-TYPE.
 FT DNA_BIND 2952 3011 HOMEBOX 4.

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP PARTIAL SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=98198453; PubMed=9531534;
 RA Steneberg P., Englund C., Kronhamn J., Weaver T.A., Samakovlis C.;
 RT "Translational readthrough in the *hdc* mRNA generates a novel branching
 RT inhibitor in the *Drosophila* trachea.";
 RL Genes Dev. 12:956-967(1998).
 CC -!- FUNCTION: Required for imaginal cell differentiation, may be
 CC involved in hormonal responsiveness during metamorphosis. Involved
 CC in an inhibitory signaling mechanism to determine the number of
 CC cells that will form unicellular sprouts in the trachea. Regulated
 CC by transcription factor *esg*. The longer *hdc* protein is completely
 CC functional and the shorter protein carries some function.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: Expressed in all imaginal cells of the embryo
 CC and larvae. Expressed in a subset of tracheal fusion cells from
 CC stage 14 to the end of embryogenesis in metameres 2-9, lateral
 CC trunk and ventral anastomoses.

```
CC      -!- MISCELLANEOUS: Readthrough of the terminator UAA occurs between
CC      codons for Ala-650 and His-652. Readthrough is not always
CC      suppressed as the shorter protein is more abundant.
CC      -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC      gene model prediction.
```

DR EMBL; Z50097; CAA90425.1; -.
DR EMBL; Z50097; CAB58233.1; -.
DR EMBL; AE003773; AAF57033.1; ALT_SEQ.
DR FlyBase; FBgn0010113; hdc.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0007430; P:terminal branching of trachea, cytoplasmic . . .; NAS.

Query Match 52.2%; Score 194; DB 1; Length 1080;
Best Local Similarity 73.2%; Pred. No. 6.5e-09;
Matches 41; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Search completed: March 12, 2004, 15:39:07
Job time : 9.94118 secs